

Figurene som det henvises til i artikkelen er gjengitt på side 26 og 27.

Vurderingspraksis i fremmedspråk

Ingrid Amundsplass, Ringerike videregående skole

27 praktiserende tysklærere i videregående skole vurderer én og samme tekst og ender opp med et sprik i karakter fra 2+ til 4/5. Én forklarer feilene, en annen setter streker, en tredje bruker farger, og en fjerde gir karakter uten begrunnelse. Med hvilken hensikt vurderer vi, og følger vi kravet om en læringsfremmende vurdering? Blir det vurdert i overensstemmelse med læreplanens kompetansemål, og hva slags kriterier legger vi vekt på?

Med disse to problemstillingene som utgangspunkt fullførte jeg våren 2011 masteroppgaven *Bewertung im Deutschunterricht* – Vurdering i tyskundervisningen – som avslutning på mastergraden «Fremmedspråk i skolen» ved Høgskolen i Østfold. I et forprosjekt to år tidligere hadde jeg oppdaget at vurderingspraksisen blant tysklærerne og ikke minst hva de la vekt på i tilbakemeldingene til elevene, varierte stort, og jeg ville derfor følge dette videre.

I teoridelen satte jeg meg inn i styringsdokumenter, læreplan og aktuell faglitteratur. I forskrift til opplæringsloven kapittel 3 står det at elevene har rett til vurdering, og at denne innebærer både rett til undervegsvurdering og sluttvurdering: «Formålet med vurdering i fag er å fremje læring underveis og uttrykke kompetansen til eleven.» Vurdering skal også «gi god tilbakemelding og rettledning til elevane [...]» (Lovdata, 2016). I følge paragraf 3.3 er det kompetanse-

målene i faget som skal utgjøre grunnlaget for vurderingen. Både med utgangspunkt i egen analyse av læreplanen og faglitteratur om vurderingsmåter valgte jeg å dele de tre hovedområdene *Språklæring, Kommunikasjon og Språk, kultur og samfunn* inn i seks kategorier som jeg brukte som vurderingskriterier i mitt videre arbeid: Innhold, Tekstforståelse, Språkbruk, Tekstbinding, Grammatikk og Språklæring (se figur 2).



Med de seks kriteriene analyserte jeg deretter en autentisk elevbesvarelse som ble brukt som halvdagsprøve på nivå II. Den var identisk med eksamenssettet *Sommerferien der anderen Art* til sentralt gitt skriftlig eksamen våren 2008. Oppgavene 1 og 2 var knyttet til en vedlagt tekst og skulle først og fremst måle tekstforståelse, oppgave 3 var et formelt brev, og de to siste var tradisjonelle skriveoppgaver. Etter en grundig analyse og oppstilling på ordnivå, konkluderte jeg med at besvarelsen hadde middels grad av målloppnåelse. Den hadde flere formelle feil, men den viste også gode kvaliteter innenfor alle de seks kriteriene. Teksten kommuniserte dessuten godt, noe fem tysktalende kunne bekrefte, og elevteksten ble bevisst valgt nettopp fordi muligheten for avvik i vurderingen var sannsynlig. 27 praktiserende tysklærere svarte på oppfordringen min og vurderte elevbesvarelsen. De fikk instruks om å gi eleven en underveisvurdering med veiledende karakter og ble også bedt om å sette en slutt karakter dersom det hadde vært en reell eksamenssituasjon. Informantene tilhørte sju ulike fylker og utgjorde en svært heterogen gruppe (se figur 1). Figuren viser også karakterene de satte.

Hva legger vi vekt på? I min analyse av elevbesvarelsen laget jeg tabeller over alle kvaliteter og formelle feil. De samme tabellene ble benyttet for registrering av markeringer eller kommentarer i informantenes korrekturer. Figur 2 og 3 viser hvor mange av lærerne som enten markerte, kommenterte eller på én eller annen måte synliggjorde for elevene at de hadde fått til eller hadde manglet når

det gjelder de nevnte kriteriene. Figur 4 viser konkrete eksempler på verbfeil og antallet som markerte hver av disse.

I figur 2 viser prosentandelen entydig at noen kriterier ble lagt mer vekt på enn andre. At Grammatikk fikk hundre prosent uttelling og at også Tekstforståelse og Språkbruk kom høyt opp, var ingen overraskelse. Oppgave 1 var spesifikt myntet på tekstforståelse, og oppgave 3 skulle være et formelt brev. Men at både Språklæring med kun 11% og Tekstbinding med 56% fikk så lite fokus, var oppsiktsvekkende. Minst like interessant som å registrere de ulike feilene, var det å se hvordan lærerne hadde synliggjort vurderingen for elevene. Noen begrunnet overhodet ikke hvorfor de valgte å sette en bestemt karakter, andre rettet opp feilene for elevene uten å gi dem muligheten til selv å komme fram til en løsning, enkelte hadde fargekoding for ulike grammatiske kategorier, og én besvarelse hadde kun streker og ikke en eneste bokstav til forklaring. Dersom dette hadde vært en reell vurderingssituasjon, ville sistnevnte tilbakemelding ha vært langt fra en oppfyllelse av styringsdokumentene som sier at vurdering skal fremme elevenes læring.

Kan vi se noe mønster i hvem som fokuserte på hva? Sett bort fra kriteriet Grammatikk, markerte hver informant cirka halvparten av kriteriene. Menn og kvinner markerte gjennomsnittlig ganske likt når det gjelder antall kriterier, men hva gjelder hvilke kriterier var det avvik: Innhold ble prosentvis markert av dobbelt så mange kvinner som menn, mens på Tekstforståelse markerte en overvekt

av kvinner på 20%. For Språkbruk sin del var menn i et knapt flertall (6%), mens for Tekstbinding synes derimot kvinnene å ha et litt sterkere fokus (12%). Ved Språklæring er ulikhetene mellom kjønnene større. 33% menn mot 10% kvinner markerte dette. Dermed virker Innhold og Tekstforståelse å ha stått mer sentralt hos kvinnene, mens Språkbruk og Språklæring var viktigere for mennene. Innenfor Grammatikk fokuserte menn og kvinner nokså likt.

I forhold til alder og erfaring ble det oppdaget en interessant forskjell i forbindelse med kriteriet Tekstbinding. I aldersgruppene over 50 var det totalt cirka 35% som fokuserte på dette, mens det til sammenligning var 90% i gruppene under 50. Tidspunktet for utdanning ser også ut til å ha spilt en rolle. Med unntak av én av de som har utdanning etter Kunnskapsløftet, markerte alle gjennomsnittlig 4 av 5 kriterier, noe som ligger langt over snittet. For kriteriet Grammatikk derimot ligger snittet litt lavere med 46% sammenlignet med 57% for alle.

/// På forhånd hadde jeg forventet at en «streng» retting med fokus på mange feil også ville føre til en lavere karakter. Dette viste seg ikke å stemme





På forhånd hadde jeg forventet at en «streng» retting med fokus på mange feil også ville føre til en lavere karakter. Dette viste seg ikke å stemme. De fire informantene som ga karakteren 2 lå på gjennomsnittet i antall kommenterte kriterier av 5, men markerte færre grammatiske feil enn gjennomsnittet for dette kriteriet. Til sammenligning hadde de med karakteren 4 samme gjennomsnitt i antall markerte kriterier, men en noe høyere frekvens på grammatikk. Eksempelvis korrigererte informantene **15** og **18** forholdsvis få grammatiske feil, men endte likevel opp på ulike karakter, 4- (3) og 2+ (2). Blant informantene som ga laveste slutt karakter befant det seg to kvinner og to menn. Av de sju informantene som ga 4 som slutt karakter var det seks kvinner og én mann. Antallet menn som deltok er for lavt til å trekke noen slutning, men kanskje viser resultatet at kvinner er noe mer opptatt av kvalitet i tekst framfor å se etter feil?

Erfaringer og nye spørsmål. Fordi de seks kriteriene var utarbeidet med utgangspunkt i læreplanen og informantene hadde tilgang til dem i sin vurdering, forutsatte jeg at alle kjente til dem og hadde tatt høyde for dem. Om de faktisk bevisst tok hensyn til dem, kunne jeg ikke bevise. Det resultatene imidlertid entydig viste var at de på én eller annen måte markerte kriteriene, og dermed ville gjort eleven oppmerksom på dem i en reell vurderingssituasjon. Lærerne markerte ikke alle kriteriene. Vekten på grammatikk var utvilsomt dominerende. Alle informantene markerte grammatiske feil på ett eller annet vis mens et knapt flertall kommenterte tekstbinding

(jf. figur 2). Evnen til å kommunisere ble altså mindre vektlagt enn språklige strukturer. Jeg konkluderte derfor med at grammatiske ferdigheter i tradisjonell forstand fortsatt er mest avgjørende i karaktersetningen. Begrunnelsen til én av informantene poengterte nettopp denne antakelsen: «Kommunikasjonen er bra i denne besvarelsen. Oppgava er forstått. Språkfeilene gjør at dette blir en middels karakter». Elevens kompetanse med tanke på tekstforståelse og kommunikasjonsevne ble altså positivt vurdert. Den virket likevel å være lite vektlagt fordi den språklige prestasjonen svekket helhetsinntrykket og dermed tross alt styrte karakteren.

Vekting av de ulike kriteriene kan og bør diskuteres. For det første varierer den fra informant til informant, og for det andre er kriterienes relevans ulik fra en situasjon til en annen. For hva ønsker vi at elevene skal få ut av fremmedspråket, og hva bør vi fokusere mest på? Forskrift til opplæringsloven har som nevnt fastsatt hensikten med vurdering. Først og fremst dreier det seg om en kontinuerlig underveisvurdering som skal motivere til videre læring. For det andre skal den være et uttrykk for elevenes kompetanse. I følge min analyse av elevteksten stemmer den gjennomsnittlige karaktergivingen til informantene (middels grad av måloppnåelse) overens med den faktiske kompetansen til eleven. Karakterforskjellene fra 2+ til 4/5 forteller for øvrig at den enkeltes oppfatning er svært varierende. Fordi det ikke ble oppdaget en klar sammenheng mellom karakterene og informantenes kjønn, alder, erfaring eller utdanning, kan en

anta at karakterresultatene er relativt representative for alle lærere. Det betyr at en kan forvente et visst avvik også ved sentralt gitt skriftlig eksamen.

I følge Bachman og Palmer (1996) er vurdering en såkalt «high-stake-decision», som påvirker hver enkelt elev og dens framtid. Sluttvurdering bør dermed på ingen måte tas for lett på. Karakterene må uttrykke elevenes reelle kompetanse og ikke være prisgitt hver enkelt lærers subjektive mening. «Kompetanse skal gi uttelling uansett hvor og hvordan den kommer til syne», skriver Utdanningsdirektoratet i veiledningsheftet til eksamen (2016b). Den reelle kompetansen til eleven i min studie viste seg imidlertid ikke å være så avgjørende for informantene, selv om kompetansemålene etter spør hva elevene kan og ikke hva de ikke kan. Fokuset på feil var svært dominerende på tross av at elevteksten hadde en større andel riktig enn feil grammatikk. Denne kvaliteten var det ingen av informantene som pekte på, og jeg fant bare noen få eksempler på positivt markerte enkeltfraser.

I et eldre rundskriv om individuell vurdering heter det at vurderingsgrunnlaget må være likt og tydelig for alle, mens det i informasjonen til elevene som følger eksamensoppgavene heter at det er «kompetansen du viser i hele svaret, som blir vurdert» (Udir, 2016a). Kompetansemålene er etter min mening ikke et tydelig nok grunnlag, med det som resultat at elevene heller ikke blir vurdert etter felles kriterier. Avslutningsvis i masteroppgaven konkluderte jeg med at validiteten og reliabiliteten på undersøkel-



sen var vanskelig å vurdere. Det skyldtes imidlertid hverken eleven selv, kompetansen hans, antallet informanter eller mitt valg av metode, men den kjensgjerning at åpne oppgaver inviterer til individuell tolking. Vil ikke det såkalte helhetsinntrykket alltid være basert på en liten eller stor porsjon skjønn?

/// Vil ikke det såkalte helhetsinntrykket alltid være basert på en liten eller stor porsjon skjønn?

Hva nå? «Avviket i resultatene av karakterer, både som sluttvurdering og undervisvurdering, avdekker behovet for mer konkrete vurderingsverktøy og kompetansemål», skrev jeg avslutningsvis i

masteroppgaven min. Nå seks år etter har jeg fått et litt annet syn på vurdering, først og fremst fordi jeg de siste tre årene har jobbet som sensor ved sentralt gitt skriftlig eksamen, men også fordi det har kommet noen eksempler på konkretisering av kjennetegn på måloppnåelse. Det som tidligere ble formulert slik: «Eleven kan [...] skrive tekster med [...] målrettet bruk av varierte språklige strukturer [...]» (Udir, 2010) er endret til «Eleven [...] bruker en variert og mer kompleks setningsoppbygging [...]» (Udir, 2016b). På sensorskolering i møte med andre tysklærere har jeg følt at vi jobber sammen for å oppnå en rettferdig vurdering i tråd med forskrifter og nasjonale kjennetegn på måloppnåelse. Vi streber etter å finne kvalitetene i elevbesvarelsene, og det er faktisk slik at fokus på grammatiske feil taper terreng dersom tekstene imponerer oss med eksempelvis godt ordforråd, innsikt i språk og kultur, eller

hvis de framstår uten kommunikasjonsforstyrrende feil. Men vi opplever også uenighet og motstridende meninger, og kanskje merker vi først og fremst frustrasjon knyttet til vage oppgaveinstruksjoner som for eksempel at elevene skal skrive en «lengre» tekst – for hva er nå det? En annen utfordring er utstrakt bruk av hjelpemidler, inkludert bruk av internett. Mitt inntrykk er at mange elever sliter med å utnytte disse «formålstjenlige», slik kjennetegn på måloppnåelse sier at de skal. Når alle hjelpemidler er tillatt, bør vel også manglende bruk av disse synliggjøres i undervisvurderingen? I følge eksamensveiledningen skal elevene i løpet av opplæringen «ha blitt veiledet i å vurdere hvilke hjelpemidler de vil ha nytte av i arbeidet med å løse ulike typer oppgaver. Det er eleven selv, gjerne i samråd med læreren, som finner frem til hvilke hjelpemidler som er formålstjenlige» (Udir, 2016b). Her har vi språklærere kanskje en vei å gå?

I skriveprosessen rundt masteroppgaven og i undersøkelsene mine fikk grammatikken mye oppmerksomhet – både av informantene og av meg. Fordi problemstillingene mine spurte etter kriterier og hensikter, dukker spørsmålet om hvor mye en skal eller bør legge vekt på grammatiske ferdigheter hos elevene opp. Hva er viktigst? Riktig form av artiklene «ein», «einen», «einem», eller at teksten som helhet kommuniserer godt og blir forstått? Under sensorskoleringen har vi fått høre at vi må legge fra oss «kjeppestene». Overdrevent fokus på grammatikk tror jeg er et treffende eksempel i så måte. Hvis målet med fremmedspråk-



undervisningen er å oppnå en best mulig karakter, så ja, da spiller all grammatikk en betydelig rolle. Men hvis målet er å ta i bruk språkkompetansen sin også etter avsluttet opplæring, betyr korrekt bruk av enkelte grammatiske former utvilsomt mindre, og vi kan ikke tillate oss å tillegge det mer vekt enn elevenes evne til å kommunisere. Germanisten Theo Harden (2006) uttaler noe slikt som at «En god karakter, både på et geografiarbeid og i en språktest, viser at eleven behersker det stoffet som er presentert i undervisningen. Etter geografiarbeidet finnes det imidlertid ikke så mye mer, mens det er etter språktesten at livet først kan begynne».

Referanser

Bachman, L. F. & Palmer, A. S. (1996): *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.

Harden, T. (2006): *Angewandte Linguistik und Fremdsprachendidaktik*. Tübingen: Narr. Egen oversettelse av sitat fra s.36.

Lovdata (2016): *Forskrift til opplæringsloven*.

Kunnskapsdepartementet. (2006). Læreplan i fremmedspråk. I: *Læreplanverket for Kunnskapsløftet (LK06)*.

Oslo: Utdanningsdirektoratet.

Utdanningsdirektoratet (2010): *Vurderingsveiledning. Fremmedspråk*.

Utdanningsdirektoratet (2016a): *Eksamen. 23.05.16 (Tysk II)*.

Utdanningsdirektoratet (2016b): *Eksamensveiledning – om vurdering av eksamensbesvarelser. Fremmedspråk*.

Figur 1: Sammenheng mellom informant, fokus på kriterier og karaktersetting

Kandidat (etter alder)	Kjønn	Alder	Praksis	ECTS (studiepoeng)	Utdanning etter K06	Sensorerfaring	Innhold	Teistforståelse	Språkbruk	Tekstbinding	Språkløring	Grammatikk (korrekturfrekvens i %)	Karakter som underveisvurdering
7	M	31-40	1	90	Ja	-	-	x	x	x	-	64	4+
10	M	31-40	10	120	-	-	x	x	x	x	-	63	3/4
17	F	31-40	3	60	Ja	-	x	x	x	x	x	48	2+
19	F	31-40	15	90	-	-	x	-	-	x	-	71	4-3
20	F	31-40	6	120	-	-	-	x	x	x	-	50	4/5
21	F	31-40	11	90	-	-	-	-	x	x	-	49	4/5
26	F	31-40	11	210	Ja	Ja	x	x	x	x	-	61	3+
27	F	31-40	7	90	-	-	-	x	x	-	-	54	4-
15	F	41-50	4	60	Ja	-	x	x	x	x	-	14	4-
25	F	41-50	17	210	-	Ja	x	x	x	x	-	55	4-
1	F	51-60	14	210	-	-	x	x	-	x	-	66	3
2	F	51-60	35	60	Ja	-	x	x	x	x	-	53	3+
4	F	51-60	25	210	-	Ja	x	x	x	x	-	52	2+
6	F	51-60	12	90	-	-	x	x	-	x	-	75	3+
8	F	51-60	32	75	-	-	-	x	x	-	-	74	3+
9	F	51-60	33	90	-	-	x	x	x	x	-	68	3+
12	F	51-60	10	130	Ja	Ja	-	-	-	-	-	39	3+
13	F	51-60	28	60	-	-	-	x	-	-	-	71	4
14	F	51-60	29	90	-	-	-	x	x	-	-	64	4
18	M	51-60	30	210	-	-	-	-	x	-	x	32	2+
23	F	51-60	25	90	-	Ja	-	x	x	-	x	40	3/4
3	F	60+	30	210	-	-	x	x	x	x	-	68	3+
5	F	60+	40	90	-	-	x	x	x	-	-	76	3+
11	F	60+	31	90	-	-	x	x	x	-	-	75	4/3
16	M	60+	38	90	-	-	-	x	-	-	-	52	2+
22	M	60+	40	60	-	-	-	x	x	-	-	47	3-
24	M	60+	33	210	-	Ja	x	-	x	x	x	67	3+

Figur 2: Forekomst av eksplisitte kommentarer til kriteriene

Kriterier	Antall (av 27)	Antall i prosent
Innhold (Oppg. 3,4,5)	15	56%
a) Svar på oppgava	12	
b) Lengde/omfang	3	
c) Sjanger	4	
d) Tittel	5	
Tekstforståelse (Oppg. 1,2)	22	81%
Oppgave 1	18	
Oppgave 2	12	
Språkbruk	23	85%
Tiltaleform oppg.3	21	
Tiltaleform oppg.5	15	
Tekstbinding	16	56%
Struktur /avsnitt	2	
Evne til å kommunisere	15	
Språklæring		
Bruk av hjelpemidler	3	11%
Grammatikk (se figur 3)	27	100%

**Figur 3: Forekomst av markerte feil innen hovedkriteriet Grammatikk**

Grammatiske kategorier (oppg.3,4,5)	Antall feil i elevtekst	Gjennomsnitt antall feil per informant	Gjennomsnitt per informant i prosent
Ordvalg	7	4,5	64 %
Ortografi	13	8,6	66 %
Kjønn og tall	4	3	75 %
Kongruens	6	5,2	87 %
Pronomen	2	0,8	40 %
Verbbygging	8	7,3	91 %
Preposisjonsfraser	11	5,2	47 %
Kasusanalyse (+ oppg.1)	5	4,3 (22/27stk)	86 %
Ordstilling	8	5,5	69 %
Adjektivbygging	3	2,4	80 %

Figur 4: Forekomst av verbfeil i elevbesvarelse og antall lærere som markerte disse

	Feil i elevtekst	Antall (av 27)
Tidsform av verb	haben überreichten	24
	will mitgenommen	25
	haben teilen	22
	will gefahren	25
	hat bekommt	25
	will hilfe	26
	fährte ich mitfährt könnte	27
Kongruens/samsvarsfeil	Das Familie können	17
	Einige Sponsoren war	15
	Wir geschenkt	24
	wir war	24
	Wir will	22
	Die Jugend haben	18
	Tanja saget ich weiß	25 27