

Automatisk rettet eksamen i generell kjemi

Maarten T. P. Beerepoot^{1*} & J. Aleksii Kosonen¹

¹Institutt for kjemi, Fakultet for naturvitenskap og teknologi, UiT Norges arktiske universitet, Norge

*Kontakt: maarten.beerepoot@uit.no

Sammendrag: En god avsluttende eksamen vurderer i hvilken grad intendert læringsutbytte har blitt oppnådd på en objektiv, reproducerbar og helst autentisk måte. Det er imidlertid lite tilfredsstillende hvis en faglærer bruker mye tid på vurdering av læring framfor vurdering *for* læring underveis i semesteret. Derfor kan det være attraktivt å utforske muligheter med en digital automatisk rettet eksamen, spesielt i emner med mange studenter. I dette bidraget beskriver vi hvorfor og hvordan vi har lagt om en eksamen i et emne i generell kjemi fra en sekstimers bruk-og-kast-eksamen på papir til en tretimers gjenbrukbar automatisk rettet digital eksamen. Vi drøfter også fordeler og ulemper med å bruke en slik eksamen som skoleeksamen eller hjemmeeksamen, systematisk kvalitetssikring av eksamensoppgavene, spørsmål knyttet til sensorens rolle i sensur av en automatisk rettet eksamen, graden av autentisitet for en slik eksamensform og avslutningsvis også overførbarhet til andre fag. Selv om en automatisk rettet eksamen ikke vil være den beste løsningen i alle fag og på alle nivåer, så mener vi at den kan være verdt å vurdere som eksamensform i grunnleggende realfaglige emner med mange studenter.

Nøkkelord:

Digital eksamen, kjemi, hjemmeeksamen, skoleeksamen, autentisk vurdering

1 Introduksjon

En god avsluttende eksamen vurderer i hvilken grad intendert læringsutbytte har blitt oppnådd på en objektiv, reproducerbar og helst autentisk måte. Et ønske om å bruke mest mulig av faglærerens tid på vurdering *for* læring kan gjøre det attraktivt å utforske muligheter med en digital automatisk rettet eksamen, spesielt i emner med mange studenter. En slik eksamen stiller ekstra høye krav til kvalitet på spørsmålene og kan gi utfordringer med for eksempel fusk og overfladisk læring blant studentene (Denyer & Hancock, 2002). Disse utfordringene er overkommelige og det finnes gode hjelpemidler til å lage gode spørsmål (Gierl et al., 2017; Gottlieb et al., 2023; Haladyna et al., 2002; Towns, 2014) og for å unngå ureglementert samarbeid og bruk av hjelpemidler (Butler-Henderson & Crawford, 2020; Nguyen et al., 2020; Rudolph et al., 2023; Sindre, 2021).

Eksamenen i KJE-1001 Introduksjon til kjemi og kjemisk biologi gjennomføres medio desember og har lenge vært en sekstimers skriftlig skoleeksamen på papir. Eksamenen besto av 20 deloppgaver der hver deloppgave ble manuelt rettet av en faglærer og skåret på en skala fra null til fem poeng med hjelp av en sensorveiledning med løsningsforslag. Prosessen med skriving og kvalitetssikring av eksamensoppgavene startet på nytt hvert år. Flere personer bidro vanligvis med oppgaver og én faglærer forsøkte å sette sammen en eksamen som var balansert, feilfri og med cirka samme vanskelighetsgrad som i tidligere år. Studentene fikk ikke lov å bruke boka eller andre hjelpemidler under eksamen, med unntak av et ark med egne notater, en kalkulator og periodesystemet. Selv om skoleeksamenen varte i seks timer brukte de aller fleste studenter mye mindre tid. Eksamenssensur ble vanligvis utført av tre faglærere rett før, i og rett etter juleferien slik at studentene først fikk eksamensresultatet tre til fire uker etter eksamen. Et mindretall av studentene ba om begrunnelse av eksamensresultatet, men de aller fleste fikk i praksis bare sluttkarakter som tilbakemelding. Tidligere eksamensoppgaver med løsningsforslag ble gjort tilgjengelig til studentene for øvelse.

Høsten 2020 ble eksamensformen lagt om til en digital automatisk rettet eksamen på tre timer. Følgende prinsipper var førende i utvikling av den nye eksamenen:

1. Faglæreren skal kunne prioritere vurdering *for* læring på bekostning av utvikling og sensur av en bruk-og-kast-eksamen. Sensur av eksamen og påfølgende tilbakemelding om resultat til studentene skal ikke ta mye tid.
2. Eksamenen skal være forutsigbar, rett på læringsmålene i emnet og samstemt med undervisning og vurdering i løpet av semesteret slik at eksamensformen oppfordrer til øvelse og læring.
3. Det skal være lov å bruke hjelpemidler på eksamen akkurat som i testene i løpet av semesteret. Dette i motsetning til en digital eksamen uten hjelpemidler som er vanlig i Norge (Raaheim et al., 2019).
4. Studentene må vise at de mestrer *alle* intenderte læringsmål for å få adgang til eksamen. Studentene som har bestått arbeidskravene uten fusk skal kunne bestå eksamenen.

I likhet med andre typer digital eksamen der hjelpemidler er tillatt (Williams & Wong, 2009) kan denne eksamenen gis både som hjemmeeksamen og som (digital) skoleeksamen.

Målet med dette bidraget er å beskrive den nye eksamensformen, samt å drøfte fordeler og ulemper med å bruke en slik eksamen som skoleeksamen eller hjemmeeksamen, systematisk kvalitetssikring av eksamensoppgavene, spørsmål knyttet til sensorens rolle i sensur av en automatisk rettet eksamen, graden av autentisitet for en slik eksamensform og avslutningsvis også overførbarhet til andre fag.

2 Metode

2.1 Beskrivelse av emnet

KJE-1001 Introduksjon til kjemi og kjemisk biologi (10 studiepoeng) er et obligatorisk emne i cirka tolv studieprogram ved fire ulike fakulteter ved UiT Norges arktiske universitet. I perioden 2020-2022 var antall kandidater på eksamen henholdsvis 221, 196 og 179 med de fleste kandidater fra studieprogrammene farmasi, biologi, geologi, biomedisin og bioteknologi. Emnet gis på høsten og de fleste studenter tar emnet i sitt første semester. Emneinnholdet er fordelt over 14 moduler/uker med fire til seks læringsmål per modul. Hver uke har studentene to til tre timer med pollbasert fellesundervisning (Micheron & Beerepoot, 2024) og to timer med seminarundervisning med gruppediskusjon basert på begrepsoppgaver (Netland et al., 2018). For å få adgang til eksamen må studentene ha bestått tre arbeidskrav: 1) laboratoriekurs, 2) oppmøte på minst 10 av 14 seminarer og 3) minst 10 av 12 poeng på alle (én eller to) innleveringer og alle (fire eller fem) obligatoriske automatisk rettede tester. Innleveringene blir rettet manuelt og tester læringsmål som ikke kan bli vurdert gjennom automatisk rettede tester, særlig læringsmål som handler om tegning og forklaring. Studenter som tar emnet på nytt må kun bestå det tredje arbeidskravet på nytt. Emnet legger opp til bruk av dokumentert effektive læringsstrategier og studentene har blant annet tilgang til en digital øvelsestest for hver modul, der studentene får ulike spørsmål hver gang de tar testen (Beerepoot, 2022). Øvelsestestene henter spørsmål fra spørsmålsbanker som har blitt utviklet med utgangspunkt i læringsmålene i dette emnet (Beerepoot, 2023).

2.2 Beskrivelse av eksamenen

En automatisk rettet eksamen på tre timer ble innført i 2020 og besto i første omgang av 36 flervalgsoppgaver som ble hentet fra spørsmålsbanker (Beerepoot, 2023). Siden 2021 er eksamenen utformet i programmet WISEflow og består av tre oppgavetyper som alle kan rettes automatisk:

1. **Flervalgsoppgaver**, vanligvis med tre eller fire svaralternativer. Eksempel: «Bestem hvilket av følgende molekyler har høyest kokepunkt ut fra type og styrke av intermolekylære krefter» med svaralternativene for én av variantene: H_2O , H_2S og H_2Te . Oppgavene gir to poeng.
2. **Matriseoppgaver** med tre til fem rader og to til tre kolonner. Eksempel: «Bestem om følgende redoksreaksjoner er spontane prosesser» med fem redoksreaksjoner (rader) der studentene må velge enten «Spontan» eller «IKKE spontan». Oppgavene gir ett poeng per rad.
3. **Rangeringsoppgaver**, vanligvis med tre grunnstoff, ioner, molekyler, salter eller bindinger som må rangeres på en bestemt måte. Eksempel: «Ranger følgende grunnstoff etter atomstørrelse fra det minste (venstre) til det største (høyre)» med

svaralternativene for én av variantene: aluminium, silisium, svovel og natrium. Oppgavene gir to poeng.

Studentene blir opplyst om at «ubesvarte spørsmål gir 0 poeng og du får ikke trekk for feil svar» med oppfordringen «dersom du ikke vet svaret: stryk svarene som uansett ikke stemmer og gjett blant de andre svarene». Mye gjetting fører til lavere reliabilitet på eksamen – særlig på eksamener med få spørsmål (Campbell, 2015). Minuspoeng for feil svar og et høyt antall svaralternativer kan motvirke denne effekten (Campbell, 2015; Gierl et al., 2017, Haladyna et al., 2002). Anbefalingen er imidlertid å heller bruke tre svaralternativer, særlig der det er vanskelig å lage plausible distraktorer (Townes, 2014; Gierl et al., 2017; Gottlieb et al., 2023; Rodriguez, 2005). Et førende prinsipp for denne eksamenen har vært å kun bruke plausible svaralternativer i flervalgsoppgaver og matriseoppgaver.

Høsten 2021 ble eksamenen gitt som hjemmeeksamen med ti ulike versjoner som var forskjellige på alle 51 oppgaver. Noen av disse oppgavene hadde i tillegg ulike varianter innenfor samme versjon, slik at ingen studenter fikk helt identisk eksamen. Høsten 2022 ble eksamen gitt som skoleeksamen med én versjon med 54 oppgaver hvorav 33 flervalgsoppgaver, 13 matriseoppgaver og 8 rangeringsoppgaver (totalt 135 poeng). Antall oppgaver med ulike varianter innenfor én versjon hadde da blitt utvidet til cirka halvparten, som er en avgjørende faktor for at eksamenen skal være gjenbrukbar. Fra og med høsten 2023 har nesten alle oppgaver ulike varianter. De ulike variantene på en oppgave har samme formulering, men for eksempel ulike molekyler som svaralternativer i en flervalgsoppgave, ulike reaksjoner som rader i en matriseoppgave eller ulike grunnstoff som må rangeres. Vanskelighetsgraden er ideelt sett likt mellom de ulike variantene og for noen oppgaver har dette blitt undersøkt og justert med hjelp av resultater fra øvelsestester eller obligatoriske automatisk rettede tester, som diskutert i Beerepoot (2023). Bokmål- og nynorskversjonene har de samme spørsmålene med de samme spørsmålsvariantene.

I tillegg til strategien med spørsmålsvarianter for å redusere *faktisk* likhet mellom eksamenssettene ble to strategier brukt for å redusere *opplevd* likhet: svaralternativene i flervalgsoppgaver og rader i matriseoppgaver ble stokket om, og hver student fikk oppgavene i en tilfeldig rekkefølge. Selv om rekkefølge av svarene kan påvirke vanskelighetsgraden, vil denne effekten jevne seg ut over eksamenen som helhet (Tellinghuisen & Sulikowski, 2008). Randomisering av rekkefølgen av svaralternativene i flervalgsoppgaver er en anbefalt strategi (Gierl et al., 2017) og følger anbefalingen om å variere posisjon av det rette svaret blant svaralternativene, men ikke anbefalingen om å bruke en logisk eller numerisk rekkefølge der mulig (Haladyna et al., 2002). Schroeder et al. (2012) fant at rekkefølgen av eksamensoppgaven kan påvirke vanskelighetsgrad og mer spesifikt særlig at en oppgave blir mer vanskelig når den etterfølger en rekke med vanskelige oppgaver. Dette har ikke blitt korrigert for eller tatt hensyn til i denne eksamenen. På grunn av randomisering av rekkefølgen av svaralternativer og rekkefølgen av eksamensoppgavene vil to studenter som sitter ved siden av hverandre nesten aldri få opp to like oppgaver. Det var imidlertid mulig for studentene å navigere fram og tilbake mellom oppgavene og studentene ble anbefalt å bruke den muligheten. Selv om begrensning av denne navigasjonen kan motvirke ulovlig samarbeid, bør man være forsiktig med å bruke dette virkemiddelet da det kan gi en ulempe for studenter med lesevansker (Sindre, 2021). I tillegg til nevnte strategier for å unngå samarbeid fikk studentene også en antifuskerklæring (Nguyen et al., 2020) tilpasset denne eksamenen,

som de bekreftet ved å levere besvarelsen. Det er all grunn til å tro at det er nødvendig å ta eksamensfusk seriøst i utforming av eksamener (Nguyen et al., 2020; Sindre, 2021) og for denne eksamenen var fokus på *forebygging* framfor *deteksjon* av fusk. Bruk av hjelpemidler som lærebok, egne notater, kalkulator og ressurser på læringsplattformen er tillatt for denne eksamenen.

2.3 Sensur

En generisk sensorveiledning ble utformet for eksamenen. Formålet med sensorveiledningen er å gi sensorene et grunnlag for å vurdere og begrunne besvarelser, gi kandidater et grunnlag for å forstå vurderingen (karakter) de har fått og gi studentene et verktøy for planlegging av arbeidet med emnet. Sensorveiledningen inneholder informasjon om arbeidskrav, pensum, oppbygning av eksamenen, tillate hjelpemidler, sammenheng mellom læringsutbyttebeskrivelse, læringsaktivitetene og vurdering i emnet, en tabell for konvertering fra poengsum til bokstavkarakter (Tabell 1) og informasjon om klageadgang. Sensorveiledningen inneholder ikke et løsningsforslag. Studentene har altså ikke tilgang til et løsningsforslag etter eksamenen, noe studentene er vant til fra oppgaver i seminarundervisning, øvelsestester og obligatoriske automatisk rettede tester i løpet av semesteret.

Tabell 1. Konverteringstabell fra poengsum til karakter for eksamen i KJE-1001 høsten 2022.

Antall poeng	Prosent	Karakter
130-135	≥ 96%	A
115-129	≥ 85%	B
95-114	≥ 70%	C
81-94	≥ 60%	D
68-80	≥ 50%	E
0-67	< 50%	F

Et utgangspunkt for konverteringstabellen (Tabell 1), som også blir kommunisert til studentene, er at minimumskravet for å bestå innleveringer og automatisk rettede tester i løpet av semesteret (10 av 12 poeng, som tilsvarer 83%) gir karakter C på eksamenen. Halvparten av poengsummen er imidlertid nok for å bestå eksamenen. Med andre ord er kravet for å bestå arbeidskravene bevisst betydelig høyere enn kravet for å bestå eksamenen. Det er imidlertid viktig å være oppmerksom på at vanskelighetsgrad for oppgavene i arbeidskravet ikke nødvendigvis er lik vanskelighetsgrad for eksamensoppgavene. Mer generelt er vanskelighetsgraden på en oppgave avhengig av ulike faktorer som for eksempel antall algoritmiske steg i en utregning, bruk av kjemisk nomenklatur framfor molekylformel og hvilke svaralternativer som er gitt i flervalgsoppgaver (Hartman & Lin, 2011). I tillegg har det noe å si i hvilken grad studentene har kunnet øve med lignende oppgaver i løpet av semesteret. Generelle utsagn om konvertering fra poeng til karakter gir altså liten mening utenfor en spesifikk kontekst. En ekstern sensor har deltatt i utforming av eksamen og sensorveiledning. Den interne sensoren har muligheten til å se på studentenes svar før poengsummen blir konvertert til karakter.

2.4 Spørreundersøkelse: studentenes perspektiver på eksamensform

Studentenes perspektiver på valget mellom skole- og hjemmeeksamen ble undersøkt i en digital avsluttende kursevaluering i november 2022. Spørreundersøkelsen inneholdt spørsmål om ulike aspekter av emnet og ble gjennomført i siste undervisningstime, og den ble i tillegg sendt ut til studenter som ikke var til stede. I en innledende tekst til delen om eksamenen ble det tydelig spesifisert at spørreundersøkelsen ikke hadde noen innvirkning på eksamensformen høsten 2022, men at eksamensformen for høsten 2023 ikke var avklart ennå. Ett lukket spørsmål med mulighet for utdypning handlet om eksamen: «Ville du foretrekke hjemmeeksamen eller skoleeksamen om du kunne velge?» med tre svarmuligheter: «jeg ville foretrekke skoleeksamen», «jeg ville foretrekke hjemmeeksamen» og «vet ikke / nøytral». For å samle en begrunnelse for valget fikk studentene som valgte skoleeksamen (hjemmeeksamen) mulighet til å utdype hva de ser som *fordeler* med skoleeksamen (hjemmeeksamen) og hva de ser som *ulempes* med hjemmeeksamen (skoleeksamen). Studentene som svarte «vet ikke / nøytral» fikk også mulighet til å utdype sitt svar. Svarene (115 utdypninger fra 104 besvarelser) ble kategorisert ut fra argumentene som ble brukt. Førsteforfatteren valgte kategoriene etter nøye gjennomlesing av alle utdypninger. Begge forfatterne telte antall utsagn i hver kategori uavhengig av hverandre og diskuterte avvik fram til enighet. Utsagn med argumenter i flere kategorier ble tatt med i flere kategorier. Kategorier med færre enn fire utsagn ble forkastet.

3 Resultater

3.1 Klager på eksamenssensur

Antall klager etter eksamenssensur var i totalt 28 i de tre årene før omleggingen (2017-2019) fra en sekstimers bruk-og-kast eksamen på papir til en tretimers gjenbrukbar automatisk rettet digital eksamen, og 0 for hvert av de tre årene etter omleggingen (2020-2022). Klager fører til en ny runde med sensur ut fra samme sensorveiledning. Litt under halvparten av klagen i perioden 2017-2019 (12 av 28; 43%) førte til en bedre karakter mens litt over halvparten (16 av 28; 57%) ikke førte til endring av karakteren. Det kan forventes at særlig studenter rett under poenggrensen til en bedre karakter klaget på karakteren, i tillegg til studenter som får lavere karakter enn forventet.

3.2 Skoleeksamen eller hjemmeeksamen?

Studentenes perspektiver på valget mellom skole- og hjemmeeksamen ble undersøkt *før* selve eksamenen. De fleste studentene hadde på dette tidspunktet ingen eksamenserfaring fra høyere utdanning. Resultatene representerer dermed i større grad studentenes *forventninger* enn deres *erfaringer*. Av de 104 studentene som svarte på kursevalueringen antydte nesten halvparten (N=50; 48%) at de ville foretrekke hjemmeeksamen og 30% (N=31) at de ville foretrekke skoleeksamen. Resten (N=23; 22%) svarte «vet ikke / nøytral». Til sammen var det 115 utdypninger. Utdypningene med argumenter for eller mot én av alternativene har blitt kategorisert og er presentert i Tabell 2.

Tabell 2. Argumenter for skoleeksamen og mot hjemmeeksamen fra studenter som foretrekker skoleeksamen, og argumenter for hjemmeeksamen og mot skoleeksamen fra studenter som foretrekker hjemmeeksamen. Kun kategorier med minst fire utsagn er gjengitt i tabellen.

Utsagn fra studenter som foretrekker skoleeksamen	# utsagn
Det er lettere å konsentrere seg på en skoleeksamen.	22
Det er lett å jukse på en hjemmeeksamen.	9
En skoleeksamen er lettere.	5
En skoleeksamen fører til mer øvelse og mer læring.	5
Utsagn fra studenter som foretrekker hjemmeeksamen	
Det er lettere å konsentrere seg på en hjemmeeksamen.	22
En hjemmeeksamen er mindre stressende / mer behagelig enn en skoleeksamen.	48

Interessant nok er det like mange utsagn som tyder på at studentene tror at det er lettere å konsentrere seg på en skoleeksamen som utsagn som tyder på det motsatte. Eksempler på utsagn fra studenter som tror det er lettere å konsentrere seg på en skoleeksamen er at «det er lettere å konsentrere seg i et arbeidsmiljø» og at det er «lettere å bli distraheret om man sitter hjemme». Eksempler på utsagn fra studenter som tror det er lettere å konsentrere seg på en hjemmeeksamen peker imidlertid på det motsatte, som for eksempel «mener selv det er mer ro når jeg sitter alene og kan konsentrere lettere» og «av egne erfaringer vet jeg at jeg er én som lett blir påvirket av lyder eller bevegelser som skjer i rommet. Det kan lett utgjøre en forskjell på resultatet jeg vil ende opp med på eksamen».

Det er derimot mer enighet om at tanken på en skoleeksamen oppleves som mer stressende, og at en hjemmeeksamen er «litt mer chill». Mange utsagn tyder på at studentene synes det er mer komfortabelt hjemme, for eksempel fordi «man kan sitte der man synes [det] er behagelig og gjøre situasjonen mindre stressende». Noen skriver at stress og nervøsitet fører til at man glemmer ting. Én student argumenterer mot skoleeksamen fordi vedkommende «ser på det som problematisk at man skal bli tatt ut av sitt vanlige miljø når man skal vise hva man kan». En annen foretrekker en hjemmeeksamen slik at man «slipper å se at andre svarer raskt om man står litt fast selv, som kan skape stress».

Andre argumenter for en skoleeksamen som kommer tilbake i minst fire utsagn er knyttet til vanskelighetsgrad på eksamen, juks, effekt på øvelse/læring eller kombinasjoner av disse. Utsagn som forklarer disse argumentene er for eksempel: «det er mye lettere å jukse på hjemmeeksamen og dermed lærer man ikke stoffet like bra, tror jeg»; på en skoleeksamen er det «[...] ikke noe kommunikasjon. Det gjør eksamen mer fair, da de med svakere nettverk ikke vil kunne få like mye hjelp [på en hjemmeeksamen] som de med sterkt nettverk rundt seg»; «Man øver ikke like mye da man tenker man googler og kan få hjelp» på en hjemmeeksamen og det er «lett å jukse og kommunisere med andre. Blir ikke en real karakter vis man skal kunne få svar av andre».

Én av studentene som svarte «vet ikke / nøytral» oppsummerer: «Tenker at hvor man sitter ikke er så viktig så lenge det er en eksamensform som tester hvordan en klarer å vise og bruke den kunnskapen man har. Men det vil være mye lettere å kontrollere at folk ikke samarbeider med en skoleeksamen».

4 Diskusjon

4.1 Oppsummering av fordeler og ulemper med en gjenbrukbar automatisk rettet eksamen

Utvikling av en automatisk rettet eksamen med mange spørsmålsvarianter krever en stor tidsinvestering og stiller høye krav til kvalitet på oppgavene. I tillegg er det en barriere for å ta i bruk et nytt digitalt verktøy for å endre vurderingspraksis (Raaheim et al., 2019). Man må være forberedt på ulike tekniske og administrative utfordringer når man planlegger en omlegging til en utradisjonell eksamensform.

Når man er ferdig med utviklingen er imidlertid fordelene mange. Faglæreren bruker lite eller ingen tid på forberedelse og sensur av eksamenen. Studentene får raskt tilbakemelding på eksamensresultat. Det er mulig å vurdere mange ulike læringsmål på kort tid, slik at det ikke er nødvendig å ta et utvalg av læringsmål som vurderes på eksamen. Eksamensformen kan brukes både som skoleeksamen og som hjemmeeksamen – særlig når det er innebygd variasjon i oppgavene som beskrevet i dette bidraget. Eksamensnivået er garantert likt fra år til år, og fra ordinær eksamen til kontinuasjonseksamen, og er godt egnet som standardisert eksamensform for samarbeid på tvers av campuser eller institusjoner. Sensur er helt objektiv og studentene vet dette, som kan ha ført til at ingen har klaget på eksamensresultat i de tre første årene etter omleggingen. Til slutt vil ikke eksamenen endre seg fra år til år, som muliggjør systematisk kvalitetsarbeid med eksamensoppgavene over årene.

4.2 Skoleeksamen eller hjemmeeksamen?

Studentenes perspektiver fra kursevalueringen tyder på at det er noen studenter som mener de kan konsentrere seg best hjemme, mens andre mener de kan konsentrere seg best på en skoleeksamen. Økt stress på en skoleeksamen er imidlertid noe som taler for en hjemmeeksamen, ifølge studentene i denne studien og i en studie utført av Williams og Wong (2009). Det er ikke sikkert at studentenes forventninger om en skoleeksamen stemmer overens med den faktiske situasjonen, gitt at de fleste trolig aldri hadde hatt en eksamen i høyere utdanning da de svarte på spørreundersøkelsen. Det er godt mulig at en skoleeksamen i høyere utdanning foregår med langt færre «lyder og bevegelser» enn det som noen studenter forestiller seg på forhånd. En begrensning av de presenterte resultatene er altså at de er preget av *forventninger* til eksamen og muligens tidligere erfaringer, heller enn å være basert på faktiske *erfaringer* med eksamen i høyere utdanning. Videre er disse resultatene ikke spesifikt knyttet til den aktuelle eksamensformen, nemlig en automatisk rettet eksamen. Det kunne være interessant å undersøke studentenes erfaringer og perspektiver *etter* eksamenen, for eksempel opplevelsen rundt eventuelt tidspress i en eksamen som tester mange ulike læringsmål på noen få timer. Selv om studentenes perspektiver er verdifulle å ta med burde disse ikke være utgangspunkt for valg av eksamensformen.

Det er relativt få studenter som argumenterer for en skoleeksamen fordi det fører til mer læring og fordi det er lett å jukse på en hjemmeeksamen (**Tabell 2**), men begge er viktige argumenter for en faglærer. Særlig rettferdighetsprinsippet og forebygging av fusk er viktig å tenke på. Sindre (2021) deler inn fusk i tre kategorier: alenefusk (for eksempel

bruk av ulovlige hjelpemidler), samarbeid (mellom kandidater som tar eksamenen) og hjelp fra en tredjeperson (som ikke tar eksamenen samtidig). Alenefusk er ikke et stort problem for denne eksamenen fordi hjelpemidler er tillatt, selv om kunstig intelligens som ChatGPT (Ali et al., 2024; Rudolph et al., 2023) har gjort det nødvendig med en begrensning til kun fysiske hjelpemidler (bøker, notater) fra høsten 2023 av. Både samarbeid og hjelp fra en tredjeperson er betydelig lettere å motvirke – men ikke helt mulig å utelukke – på en skoleeksamen der studentene er under observasjon (Butler-Henderson & Crawford, 2020). Selv om hjelp fra en tredjeperson absolutt ikke kan utelukkes verken i arbeidskrav eller på eksamen, så er dette et mindre problem i et emne som dette der en student må bestå både eksamen og ulike tester, innleveringer og andre arbeidskrav gjennom semesteret (Williams & Wong, 2009). Samarbeid er i praksis umulig å unngå på en hjemmeeksamen og er sannsynligvis det største problemet for denne eksamenen. Selv om antifuskerklæringen, de ulike spørsmålsvariantene og den tilfeldige rekkefølgen av spørsmål og svaralternativer vil redusere samarbeid til en viss grad, vil de ikke kunne forhindre det. Økt tidspress kan gjøre at færre har mulighet å delta i samarbeid, men kan også føre til mer eksamensangst og dermed mer fusk (Sindre, 2021). Den mest effektive og rettferdige måten å begrense samarbeid på er derfor å gi eksamenen som skoleeksamen. Eksamenen i KJE-1001 vil derfor inntil videre være en digital automatisk rettet skoleeksamen.

4.3 Kvalitetssikring av eksamensoppgaver

Uansett om eksamenen blir brukt som skole- eller hjemmeeksamen, stiller en automatisk rettet eksamen enda høyere krav til kvalitet på oppgavene enn en manuelt rettet eksamen. Studentene har ingen mulighet til å begrunne sine svar og dermed er det vanskelig for en sensor å oppdage uheldige formuleringer i oppgaveteksten. Studentene har ikke innsikt i sin besvarelse etter eksamen, og dermed kan de heller ikke argumentere for sine svar i etterkant av sensur. Dette legger et stort ansvar på faglærerne som utvikler eksamenen, og det stilles høye krav til kvalitetssikring. Beerepoot (2023) har presentert fire strategier for kvalitetssikring av spørsmålsbanker som brukes i automatisk rettede tester. Noen av disse strategier har vært særlig nyttig i utviklingen av den aktuelle eksamenen. Noen av flervalgsoppgavene har blitt brukt og kvalitetssikret i *no-stake* øvelsestester og *low-stake* obligatoriske tester før de ble brukt i eksamenen. Gjennom analyse av testresultatene har eventuelle feil blitt rettet opp og dårlige spørsmål fjernet. Tre faglærere har testet hele eksamenen flere ganger, diskutert eventuelle gale svar og forbedret oppgavene i prosessen. En annen kvalitetssikringsrutine som har blitt brukt, er å gå gjennom besvarelsene fra studentene som hadde kun én eller to feil i sin besvarelse med en hypotese at det var en feil i utforming av oppgaven. Selv om hypotesen ikke pleier å være riktig, har denne strategien ført til at én feil i eksamen har blitt funnet og rettet opp i eksamenssensuren. Den store fordelen med en gjenbrukbar eksamen er at slike rutiner for kvalitetssikring fører til en systematisk forbedring av kvalitet på eksamenen gjennom årene, i motsetning til kvalitetssikring av en bruk-og-kast-eksamen der man starter på nytt hver eneste gang. En årlig gjennomgang av eksamensoppgavene etter kontinuasjonseksamenen på våren har blitt innarbeidet som rutine for systematisk kvalitetsarbeid med eksamenen i KJE-1001.

4.4 Problemstillinger knyttet til sensur

Bruk av en automatisk rettet eksamen med mange spørsmålsvarianter stiller også andre krav til rollen som ekstern sensor enn en mer tradisjonell manuelt rettet eksamen. En ekstern sensor kan ha en rolle i utforming av en eksamen. Det er imidlertid uvanlig både at en eksamen har mye variasjon fra student til student innenfor samme år, og at den har lite eller ingen variasjon fra år til år. Burde en ekstern sensor se på alle eksamensoppgaver på nytt hvis ett spørsmål blir lagt til eller byttet ut? Burde en ekstern sensor gå gjennom alle mulige varianter av et spørsmål og – hvis ja – hvordan? En ekstern sensor kan også ha en rolle i vurdering av studentenes besvarelser. Denne vurderingen gjennomføres imidlertid i utgangspunkt automatisk. Hvordan kan en ekstern sensor bidra til kvalitet av sensur? Hva gjør klagesensorene, som blir bedt om å vurdere en besvarelse på nytt ut fra samme sensorveiledning som også inneholder konverteringstabellen fra poengsum til karakter? Hvordan kan studentenes rettigheter til klage ivaretas, hvis verken den første sensoren eller klagesensoren egentlig foretar en vurdering av studentenes besvarelse? Disse er noen av spørsmålene som burde diskuteres når man bruker en automatisk rettet eksamen. Et administrativt aspekt som er verdt å se nærmere på er regelverket rundt sensur og klage for en automatisk rettet eksamen. En ekstern sensor har mer å bidra med i utformingen av oppgaver og kvalitetssikringen, enn i vurderingen av besvarelsene.

4.5 Autentisk vurdering

Er denne eksamensformen en autentisk måte å vurdere på? Villarroel et al. (2018) presenterer tre dimensjoner som sammen dekker essensen av autentisk vurdering: 1) oppgaven skal være i en realistisk kontekst, eller skal være lik en ekte problemstilling som kandidatene kan komme til å møte i arbeidslivet, 2) kandidaten skal bruke høyere kognitive ferdigheter som problemløsning for å utvikle kognitive og metakognitive ferdigheter, og 3) kandidaten skal utvikle ferdigheter for å kunne vurdere sin egen prestasjon og regulere sin egen læring. Et anbefalt første steg for å innføre autentisk vurdering er å identifisere hva en student må kunne etter å ha gått gjennom studiet, for eksempel formulert i en læringsutbyttebeskrivelse for et spesifikt studieprogram (Villarroel et al., 2018). Hvordan skal man i så fall forholde seg til den første dimensjonen om realisme i et begynneremne med studenter fra mange ulike studieprogram? Mulighetsrommet for å vurdere på en autentisk måte er altså begrenset i slike emner. I dette innføringskurset i generell kjemi – som de aller fleste studenter tar i sitt første semester – ligger ikke fokus på forberedelse på arbeidslivet, men heller på forberedelse til videre studier i høyere utdanning gjennom vektlegging av og tilrettelegging for bruk av effektive læringsstrategier (Beerepoot, 2022). Et aspekt av dette som er felles med den tredje dimensjonen i en autentisk vurderingspraksis, er å støtte studentenes selvregulering av læring med en gjennomtenkt vurderingspraksis som vektlegger formativ vurdering (Nicol & Macfarlane-Dick, 2006). Selv om noen vil mene at flervalgsoppgaver er motstridende med prinsippene for autentisk vurdering, er det behov for å nyansere dette bildet fordi at automatisk rettede tester med flervalgs-oppgaver *i løpet av semesteret* har potensial til å støtte selvregulert læring (Beerepoot, 2022, 2023; Nicol, 2007). I teorien er det mulig å introdusere en kontekst i automatisk rettede eksamensoppgaver som et bidrag til realisme og til bruk av høyere kognitive ferdigheter

(Nguyen et al., 2020; Villarroel et al., 2018). Dette strider imidlertid mot et viktig prinsipp i konstruksjon av flervalgsoppgaver, nemlig å begrense mengden informasjon til et minimum for å kunne svare på oppgaven (Gottlieb et al., 2023; Haladyna et al., 2002; Towns, 2014).

Man kan likevel argumentere for at denne omleggingen til en automatisk rettet digital eksamen har ulike aspekter som fører til en *mer* autentisk vurderingspraksis: eksamenskandidater kan bruke ulike hjelpemidler under eksamen; eksamenen gis i et digitalt format (Butler-Henderson & Crawford, 2020); kandidatene må svare på mange oppgaver innen kort tid (Butler-Henderson & Crawford, 2020); og ikke minst frigjør det tid for faglæreren til å vektlegge vurdering *for* læring framfor vurdering *av* læring.

4.6 Overføringsverdi

De fleste aspektene diskutert i dette bidraget er overførbare til andre emner og fag der en automatisk rettet eksamen vurderes som (del av) den summative vurderingen. Det gjelder for eksempel fordelene og ulempene med eksamensformen generelt (§4.1), de høye kravene til kvalitetssikring og mulighetene for *systematisk* kvalitetssikring fra år til år (§4.3) og problemstillingene knyttet til sensur av en automatisk rettet eksamen (§4.4). En automatisk rettet eksamen sikrer objektivitet og reproduserbarhet i sensuren. Ikke minst fører eksamensformen til en begrensning av faglærerens tidsbruk *etter* eksamen, som kan muliggjøre et større fokus på formativ vurdering *før* eksamen.

Eksamensformen passer imidlertid ikke til alle emner og alle fag, og spesielt ikke til alle nivåer. Eksamensformen egner seg spesielt godt til begynneremner med mange studenter og en stor bredde i læringsmål som det er ønskelig å teste på en eksamen. Dersom læringsutbyttet kan testes på en tilfredsstillende måte gjennom automatisk rettede oppgaver – gjerne i kombinasjon med arbeidskrav som passer til fagets egenart – kan det være verdt å vurdere denne eksamensformen.

Referanser

- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J. & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6), 643.
<https://doi.org/10.3390/educsci14060643>
- Beerepoot, M. T. P. (2022). Effektive læringsstrategier – Innsikter, implementering og tverrfaglig erfaringsutveksling. *Læring om læring*, 9(2), 9.
<https://www.ntnu.no/ojs/index.php/lo/article/view/5020>
- Beerepoot, M. T. P. (2023). Formative and summative automated assessment with multiple-choice question banks. *Journal of Chemical Education*, 100(8), 2947–2955.
<https://doi.org/10.1021/acs.jchemed.3c00120>
- Butler-Henderson, K. & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, 159, 104024.
<https://doi.org/10.1016/j.compedu.2020.104024>
- Campbell, M. L. (2015). Multiple-choice exams and guessing: Results from a one-year study of general chemistry tests designed to discourage guessing. *Journal of Chemical Education*, 92(7), 1194–1200.
<https://doi.org/10.1021/ed500465q>
- Denyer, G. & Hancock, D. (2002). Graded multiple choice questions: Rewarding understanding and preventing plagiarism. *Journal of Chemical Education*, 79(8), 961–964.
<https://doi.org/10.1021/ed079p961>

- Gierl, M. J., Bulut, O., Guo, Q. & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, 87(6), 1082–1116. <https://doi.org/10.3102/0034654317726529>
- Gottlieb, M., Bailitz, J., Fix, M., Shappell, E. & Wagner, M. J. (2023). Educator's blueprint: A how-to guide for developing high-quality multiple-choice questions. *AEM Education and Training*, 7(1), e10836. <https://doi.org/10.1002/aet2.10836>
- Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, 15(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5
- Hartman, J. R. & Lin, S. (2011). Analysis of student performance on multiple-choice questions in general chemistry. *Journal of Chemical Education*, 88(9), 1223–1230. <https://doi.org/10.1021/ed100133v>
- Micheron, D. & Beerepoot, M. T. P. (2024). Studentaktiv læring med store studentgrupper – flervalgsoppgaver i sentrum. *Nordic Journal of STEM Education*, 8(1), 17–26. <https://doi.org/10.5324/njsteme.v8i1.5117>
- Netland, K. Ø., Sivertsen, A. & Olufsen, M. (2018). Innføring av studentaktive arbeidsformer i seminarundervisningen. Hvilken betydning har dette på læringsutbyttet og klasse miljøet? *Nordic Journal of STEM Education*, 2(1), 15. <https://doi.org/10.5324/njsteme.v2i1.2346>
- Nguyen, J. G., Keuseman, K. J. & Humston, J. J. (2020). Minimize online cheating for online assessments during COVID-19 pandemic. *Journal of Chemical Education*, 97(9), 3429–3435. <https://doi.org/10.1021/acs.jchemed.0c00790>
- Nicol, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education* 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education*, 31(1), 53–64. <https://doi.org/10.1080/03098770601167922>
- Raaheim, A., Mathiassen, K., Moen, V., Lona, I., Gynnild, V., Bunæs, B. R. & Hasle, E. T. (2019). Digital assessment—how does it challenge local practices and national law? A Norwegian case study. *European Journal of Higher Education*, 9(2), 219–231. <https://doi.org/10.1080/21568235.2018.1541420>
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational measurement: issues and practice*, 24(2), 3–13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>
- Rudolph, J., Tan, S. & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.9>
- Schroeder, J., Murphy, K. L. & Holme, T. A. (2012). Investigating factors that influence item performance on ACS exams. *Journal of Chemical Education*, 89(3), 346–350. <https://doi.org/10.1021/ed101175f>
- Sindre, G. (2021). Kan fusk på hjemmeeksamen forhindres? *Nordic Journal of STEM Education*, 5(1), 5. <https://doi.org/10.5324/njsteme.v5i1.3918>
- Tellinghuisen, J. & Sulikowski, M. M. (2008). Does the answer order matter on multiple-choice exams? *Journal of Chemical Education*, 85(4), 572–575. <https://doi.org/10.1021/ed085p572>
- Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, 91(9), 1426–1431. <https://doi.org/10.1021/ed500076x>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C. & Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854. <https://doi.org/10.1080/02602938.2017.1412396>
- Williams, J. B. & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, 40(2), 227–236. <https://doi.org/10.1111/j.1467-8535.2008.00929.x>