Feedback 2.0 – från penna till AI

En undersökning av AI-genererad formativ bedömning av gymnasieelevers debattartiklar

Examensarbete inom ämneslärarprogrammet inriktning gymnasiet

Titel (eng): Feedback 2.0 – From Pen to AI: An Investigation of AI-Generated Formative Assessment of Debate Articles

Författare: Chanel Hallkvist

Termin och år: Vårtterminen 2024

Kurs: LGSV2A 15hp

Kursansvarig institution: Institutionen för språk och litteraturer, Göteborgs Universitet

Handledare: Ann-Christin Randahl

Examinator: Lena Rogström

Nyckelord: formativ bedömning, formativ feedback, AI-genererad feedback, skrivutveckling, ChatGPT, skrivundervisning, *AI-assessment, AI-feedback, formative feedback, formative assessment*

Abstract:

This study explores how AI-generated feedback affects high school students' writing. Twenty-four students got feedback on their debate articles from ChatGPT 3.5 and then revised their texts based on this feedback. The majority of the students' texts improved in quality after using at least two feedback comments from AI. However, some students struggled to understand the feedback fully, highlighting the need for teachers to include an explicit dialogue to elaborate the feedback. This study contributes to the limited existing research on the integration of AI in assessment and wishes to provide insights into its potential benefits and limitations in supporting students' ability to develop their texts.

Förord

Största möjliga tack till min handledare, Ann-Christin Randahl, varit min krycka, cheerleader och kritiker längs detta äventyr. Detta arbete hade inte alls fått sin form utan hennes vägledning och jag är otroligt tacksam för all konstruktiv kritik, tankar och idéer hon bidragit med. Också ett stort tack till alla informanter som gjorde denna studie möjlig att genomföra. Slutligen vill jag tacka Lena Rogström samt Ester Blomberg för er betydelsefulla input vid uppsatsventileringen.

INNEHÅLL

1 Inledning---- 5
1.1 Bakgrund----5
1.2 Syfte och frågeställningar----6
2 Tidigare forskning---- 7
2.1 Formativ feedback i skrivundervisning---- 7
2.1.1 Revideringsmönster---- 8
2.2 AI i skrivundervisning----10
3 Teoretisk inramning---- 13
3.1 Formativ bedömning---- 13
3.1.1 Teorins tillämpning----14
4 Metod och material---- 15
4.1 Datainsamling och urval---- 15
4.1.1 Insamling av texter---- 15
4.1.2 AI-bedömning---- 16
4.1.3 Gruppintervjuer---- 17
4.2 Analys av data---- 18
4.2.1 Kvalitativ innehållsanalys av AI-genererad feedback och revidering av elevtext---- 18
4.2.2 Tematisk analys av gruppintervjuer---- 22
4.3 Etiska överväganden----26
5 Resultat---- 28
5.1 Övergripande bild av ChatGPTs feedback och elevernas revidering---- 28
5.2 Fördjupad bild av elevers revidering---- 30
5.2.1 Revideringar på global nivå---- 32
5.2.2 Revideringar på lokal nivå---- 34
5.2.3 Sammanfattning av globala och lokala revideringar---- 35
5.3 Elevernas upplevelse av AI-genererad formativ feedback---- 36
5.3.1 Användbarhet----36
5.3.2 Effektivitet----37
5.3.3 Begriplighet----38
5.3.4 Tillförlitlighet----40
5.3.5 Tilltal---- 41
5.3.6 Förändring----42
5.4 Sammanfattning---- 42
6 Diskussion---- 44
6.1 Studiens resultat i relation till tidigare forskning---- 44
6.2 Didaktiska implikationer---- 45
6.2.1 Effektiv resursanvändning---- 45
6.2.2 Utmaningar med effektivitet---- 46
6.2.3 Självständigt lärande---- 46
6.2.4 Normering och kreativitet---- 46
6.3 Metodreflektion----46
6.4 Fortsatt forskning---- 47
7 Referenslista----48
8 Bilagor---- 51
Bilaga 1: Prompt inklusive deltagarnas uppgiftsbeskrivning till debattartikel---- 51
Bilaga 2: Bedömningsmatris för debattartikel Svenska 1---- 52
Bilaga 3: Exempel på AI-genererad feedback på en stark debattartikel----53
Bilaga 4: Exempel på AI-genererad feedback på en svag debattartikel----54
Bilaga 5: Intervjuguide----55

1 Inledning

1.1 Bakgrund

Den snabba teknologiska utvecklingen har gjort artificiell intelligens (AI) alltmer framträdande inom olika sektorer i dagens samhälle. I skolkontexten ser man med både oro och förhoppningar på detta nya verktyg (Hedman, 2023). De mer positiva menar att implementeringen av AI kan öppna dörren för innovativa metoder för planering, undervisning och bedömning. Enligt Fidan och Gencel (2022) kan användningen av chatbots som feedbackverktyg förbättra elevers inlärning genom att erbjuda detaljerad och omedelbar återkoppling. Enligt rapporter från World Economic Forum (2023) kommer AI spela en betydande roll i framtidens skolor. De menar att det är nödvändigt att utbildningssystemen anpassar sig till och lär sig av dessa teknologiska framsteg.

Holmes et al. (2019) och Hänze och Pinquart (2020) är eniga om att användningen av AI-baserad bedömning kan vara en resurs för både lärare och elever. De betonar att denna teknik har potential att erbjuda snabb och objektiv feedback, avlasta lärare från en omfattande bedömningsbörda samt förbättra elevernas skrivutveckling genom snabbare och mer precisa återkopplingsprocesser (Holmes et al., 2019; Hänze & Pinquart, 2020).

Enligt OpenAI (n.d.a) som skapat AI-modellen ChatGPT, finns det också utmaningar. Exempelvis är ChatGPT inte kapabel till att fullt ut förstå den komplexa karaktären hos varken en elev eller i en pedagogisk kontext. Det finns även risk att det förekommer felaktigheter och bristande objektivitet bland de AI-genererade svaren. Enligt modellens användningspolicy rekommenderas den dessutom inte att användas för bedömningsbeslut utan att en mänsklig granskande faktor finns med i processen.

Även om det ibland förs en kritisk diskussion om prov och bedömningar i skolorna är bedömningar oundvikliga och dessutom värdefulla för elevers lärande, särskilt om de är framåtsyftande eller formativa. Enligt OECD (2005) är formativa bedömningar en central strategi för att förbättra elevers prestation och engagemang i skolan. Vid formativa bedömningar identifieras specifika områden där eleven behöver stöd. Formativ bedömning ges kontinuerligt under lärandeprocessen och syftar till att vägleda elever i denna process (Jakobson, 2019). Det är här verktyg som ChatGPT kan vara särskilt gynnsam för undervisning och lärande, eftersom den är snabb och enligt OpenAI kan ge detaljerad återkoppling. Att använda digitala verktyg för bedömning kan ge lärare större möjligheter att individanpassa sin undervisning och stöd i att synliggöra elevernas framsteg.

1.2 Syfte och frågeställningar

Studiens övergripande syfte är att bidra med kunskap om effekter av formativ feedback av elevtexter producerad av artificiell intelligens (AI) inom svenskundervisningen på gymnasiet. Genom att analysera AI-bedömda elevtexter samt att utforska elevers attityder och upplevelser av denna feedback önskar studien kasta ljus över AIs potentiella användbarhet och framtida roll i skolans skrivundervisning. För att besvara syftet ställs följande fyra frågor:

- På vilka språkliga nivåer kan AI ge formativ feedback av elevtexter?
- Vilken effekt har AI-genererad formativ feedback på elevers revidering av skrivna texter?
- Hur upplever elever att få AI-genererad feedback?
- Vilka pedagogiska möjligheter och utmaningar kan identifieras vid användning av AI-feedback inom skrivundervisning?

2 Tidigare forskning

I detta kapitel presenteras tidigare forskning som är relevant för denna studie. Tre huvudområden kommer att vara i fokus: formativ feedback i skrivundervisning, elevers revideringsmönster samt användning av AI i skrivundervisning.

2.1 Formativ feedback i skrivundervisning

Redan 1968 introducerades konceptet formativ bedömning i skrivundervisningen. Det var Bloom (1968, s.9) som utvecklade begreppet "mastery learning", baserat på idéer om formativ bedömning. Blooms koncept fokuserade på att organisera undervisningen i enheter och använda regelbundna formativa bedömningar för att ge återkoppling till eleverna om deras inlärning. I Blooms modell skulle varje undervisningsenhet avslutas med en kort formativ bedömning för att bestämma om eleverna hade behärskat enhetens innehåll. Denna bedömning används inte för att betygsätta eleverna utan snarare för att ge dem information om deras framsteg och identifiera eventuella kunskapsluckor. Om eleverna visade sig behärska enheten fick de fortsatt stöd för att stärka sin kunskap. Dessa elever fortsatte då med nästa enhet med ökat självförtroende. De elever som inte behärskade innehållet fick istället individuell vägledning och stöd för att komma ikapp resten av klassen (Bloom, 1971, s.58).

I en ofta citerad studie av Hattie och Timperley (2007, s.87) betonas feedbackens avgörande roll för elevers lärande. I sin artikel har de utformat en holistisk strategi som syftar till att optimera elevers inlärning. Strategin består av tre centrala komponenter: *feed up, feed back* och *feed forward*. Komponenten Feed up, involverar att förtydliga lärandemålen och förväntningarna för eleverna och anses vara en förberedande fas som skapar en ram för kommande feedbackprocess. Feed up ska ge eleverna tydliga förväntningar på vad som krävs för att nå målen för uppgiften samt vägledning i hur eleverna kan bedöma sin egen utveckling mot dessa mål. Komponenten Feed back representerar det faktiska tillfället då återkoppling ges till eleverna baserat på deras prestation i förhållande till målen. Denna feedback ska inte enbart identifiera brister utan också framhålla styrkor och erbjuda konkreta åtgärder för förbättring av texten. Den tredje komponenten, feed forward, fokuserar på framtida strategier baserat på den feedback som eleven mottagit, med syftet att förbättra elevens framtida prestationer och främja kontinuerligt lärande. Forskning har visat att elever som ges möjlighet att reflektera över och agera på feedback kan öka sin förståelse och förbättra sina prestationer över tid (s.104).

I forskningsöversikten "Feedback i skrivundervisningen" har Skolforskningsinstitutet (2015) sammanställt forskning från 22 studier mellan årtalen 2000–2016 som utforskar feedbackprocessen inom skrivundervisningen. En central slutsats är att en god relation mellan lärare och elev är av stor betydelse. Det framgår även att elevernas uppfattning om feedbackens relevans och användbarhet är nära kopplad till deras engagemang och förmåga att utveckla sina texter. Det är därför väsentligt att tydligt förklara syftet med feedbacken och dess koppling till mål för skrivandet så att eleven kan skapa en förståelse för hur de kan förbättra sin skrivförmåga (Skolforskningsinstitutet, 2015). Enligt översikten skapas denna förståelse genom dialog. Dialogen syftar till att både lärare och elever ska förstå vilka förändringar som behöver göras och varför de är viktiga.

I en avhandling av Jakobson (2019) utforskas hur lärarrespons på skriftliga texter för vuxna nybörjare i andraspråksinlärning fungerar. Det är ett område som inte tidigare fått mycket uppmärksamhet, men det är viktigt att förstå hur lärare kan ge effektiv respons eftersom nybörjare liksom oerfarna skribenter behöver en stark grund för att utveckla sin språkkompetens. Jakobson (s.44) visar att olika typer av respons kan ha olika effekter på studenters inlärning och språkutveckling. Kategorier som *innehåll, organisation* och *språklig korrekthet* identifierades som möjliga fokuspunkter för respons. Dessa kategorier som är framtagna för att utveckla avancerade skribenters kompetens, visade sig även vara användbara för nybörjare. Jakobson identifierar tre typer av respons: beröm, kritik eller förslag till förbättringar. Läraren väljer vad de vill fokusera på i sin respons, exempelvis språklig korrekthet, vilket bestämmer inriktningen på responsen. Enligt Jakobson (s.99) är alla tre av responssätten nödvändiga. Specifikt beröm uppskattades mest av studenter, särskilt när den var kopplat till konkreta framsteg. Detta kan öka motivationen hos studenterna och hjälpa dem förstå vad de gör bra. Specifik kritik visade sig också vara uppskattat då den talade om för eleven vad hen specifikt behövde förbättra. Allmänt beröm på en text visade sig vara mindre effektivt, eftersom det inte tydligt visade inom vilka områden studenten hade gjort framsteg.

2.1.1 Revideringsmönster

I forskningsområdet kring textrevidering har det funnits ett intresse för revideringsmönster hos olika grupper av skribenter. Faigley och Witte (1981) bidrog till denna forskning genom att genomföra en undersökning av revideringsmönster hos erfarna och oerfarna skribenter. En viktig insikt som framkom i deras arbete var att erfarna skribenter tenderar att ägna sig åt en

mer djupgående revideringsprocess jämfört med vad oerfarna skribenter gör. Resultatet visade att erfarna skribenter ofta utförde makrostrukturella förändringar för att förbättra tydligheten i sina texter. Dessa revideringar syftade inte bara till att korrigera ytliga brister utan också fördjupa och utveckla textens innehåll och struktur. Oerfarna skribenter tenderade däremot att främst fokusera på stavfel och grammatik, utan att göra betydelsebärande förändringar på textens meningsnivå.

En väletablerad modell för att synliggöra olika nivåer av texten är den så kallade texttriangeln (Dysthe et al., 2011, s.41) som presenterar olika nivåer av textproduktion och revidering. Varje nivå av triangeln representerar olika aspekter av skrivandet och består av fem nivåer: skrivsituation, innehåll/struktur, meningar, ordval samt stavningskontroll/användning av skiljetecken/formalia i samband med referenser. Dessa nivåer sträcker sig från en global (högre) nivå till en lokal (lägre) nivå. Varje nivå är viktig för att skapa en sammanhängande och välskriven text. När det gäller revidering är det avgörande att förstå hur varje nivå påverkar textens kvalitet. Det finns ett samband mellan de textnivåer som revideras och den färdiga textens kvalitet. Dysthe et al. konstaterar, liksom Kronholm-Cederberg (2009) att erfarna skribenter tenderar att ägna mer tid åt att revidera på de globala nivåerna, såsom struktur och innehåll. Mindre erfarna skribenter fokuserar mer på lokala nivåer, såsom ordval och grammatik. En möjlig förklaring till detta är att erfarna skribenter har en djupare förståelse för olika genrer och lingvistiska mönster. Denna förståelse bidrar till att de kan göra mer genomgripande revideringar på de högre nivåerna. Mindre erfarna skribenter tenderar att ha en begränsad repertoar och fokuserar istället på mindre detaljer i texten (Dysthe et al., 2011, s. 53).

I sin avhandling om responskultur i förstaspråksundervisning på gymnasiet synliggör författaren Kronholm-Cederberg (2009) olika mönster i skolans responskultur. Studien visar att skolans responskultur domineras av ett bristfokus. Eleverna möts av övervägande negativ respons och känner också att deras skriftliga uttryck och åsikter bedöms normativt. Detta leder till att skrivutvecklingen kan begränsas och att oerfarna och erfarna skribenters positioner cementeras (s.106). I Kronholm-Cederbergs avhandling framträder skillnader mellan skribenters förmågor att navigera mellan globala och lokala nivåer av revidering. Goda skribenter visar en förmåga att prioritera innehållsfrågor och fokuserar på att göra texten till sin egen genom en medveten och självständig hantering av språket. Kronholm-Cederberg menar att revidering utifrån en global textnivå kräver mer komplexa kognitiva processer av skribenten då det involverar större textmängder. Dessa textnivåer är inte heller lika konkreta som de lokala (s. 205). I kontrast till detta fokuserar mindre goda

skribenter på detaljer på en mer lokal nivå, som stavning och ordval. Dessa skribenter riskerar att förlora det övergripande syftet med texten. De saknar ofta också en egen röst. Responskulturens dilemma är enligt Kronholm-Cederberg att balansera lärarens begränsade tid för återkoppling och behovet av en mer ingående och kvalitativ feedback på de globala aspekterna av elevernas texter. Att ge elever feedback på global nivå tenderar att kräva en större arbetsinsats och ett större engagemang av läraren då texten behöver granskas och analyseras som helhet. Det kräver också en djupare förståelse för ämnet och elevens intentioner med texten, vilket är mer tidskrävande än att korrigera grammatiska fel på en lokal nivå (s. 277).

2.2 AI i skrivundervisning

AI i skrivundervisningen är ett nytt forskningsområde och det finns ännu (våren 2024) begränsat med studier. I en undersökning om användningen av språkmodeller i skrivundervisning som utfördes av Zirar (2023, s.8) framgår det att språkmodeller, som är en chattrobot som genererar information på ett människoliknande vis likt ChatGPT, kan ha betydande konsekvenser för elevers lärande. Att enbart förlita sig på språkmodeller för att ta del av information kan negativt påverka studenters förmåga att tänka kritiskt. Eleven accepterar svaret från ChatGPT utan att ifrågasätta och lär sig inte skilja mellan sanningsenlig information och faktafel, menar Zirar. Språkmodeller kan ge acceptabla svar men saknar förmågan att förstå innebörden och avsikten på det uttalande den ger svar på. Det är även viktigt att komma ihåg att en acceptabel förklaring skiljer sig från en korrekt förklaring. Studenterna bör därför göras medvetna om den ibland bristfälliga tillförlitligheten i den genererade informationen. Zirar menar även att språkmodeller kan vara användbara för att generera undervisnings- och bedömningsmaterial samt utvärdera studenters svar på sådant material. Trots att detta är ett verktyg som kan producera stora mängder material och bidra till en effektiv återkopplingsprocess är det enligt Zirar viktigt att vara medveten om att människans roll fortfarande är väsentlig då materialet behöver granskas för att säkerställa dess kvalitet och tillförlitlighet.

Av en litteraturöversikt om användningen av ChatGPT i undervisning och lärande (Duong et al., 2024, s.13–14) framgår att det både finns fördelar och risker med att använda ChatGPT i undervisningen. Bland fördelarna menar Duong et al. att ChatGPT har potential att förbättra inlärningen hos elever genom interaktiva utbyten och skräddarsydda svar. Den kan även öka elevers motivation och självförtroende genom att stödja eleverna i språkinlärning

och problemlösning. ChatGPT kan, enligt Duong et al., också fungera som en slags "skrivpartner" och erbjuda omedelbar feedback och vägledning vid elevers skrivande. Verktyget kan även fungera som en resurs för att generera idéer och inspiration inför olika skrivprojekt. Eleven kan mata in ett ämne eller en fråga och få förslag på ämnesområden och strukturella idéer som kan hjälpa dem att komma igång med eller fortsätta sitt skrivande. Men författarna ser även en risk att ChatGPT ersätter snarare än stödjer elevers skrivande. Det är bland annat svårt för lärare att säkerställa att elever utvecklar sina egna skrivförmågor och inte bara förlitar sig på ChatGPT. För att elever ska använda verktyget på ett effektivt sätt är det viktigt att läraren skapar en balanserad undervisningsmiljö där elever får använda ChatGPT som ett komplement till sin skrivprocess samtidigt som de uppmanas att utveckla sin skrivkompetens. Detta kan ske genom att eleverna kritiskt reflekterar över hur de kan förbättra sina egna skrivfärdigheter genom denna interaktion**.** En annan utmaning som lyfts är risken för fusk vid examinationer samt att verktyget kan ha en negativ inverkan på både elevers förmåga att tänka kritiskt och att kommunicera (Duong et al., 2024).

En studie av Sok och Heng (2023, s.3) visar att användandet av ChatGPT spara lärare tid och ansträngning. Lärare kan semi-automatisera bedömningen av studenters arbete då ChatGPT kan identifiera både svagheter och styrkor. Sok och Heng identifierar också nackdelar, framförallt att eleven enkelt kan fuska då ChatGPT har förmågan att generera svar och texter som ibland kan vara svår att särskilja från elevers autentiska arbete.

I en översikt av Montenegro-Rueda et al. (2023, s.7) visar det sig att användningen av ChatGPT fungerar som ett betydande stöd för undervisnings- och lärandeprocessen för både elever och lärare. Samtidigt synliggör författarna att det också finns utmaningar med att effektivt implementera detta verktyg i undervisningen. Till exempel saknar lärare explicit utbildning när det gäller användning av ChatGPT i sin undervisning och därmed nödvändiga färdigheter för att använda det som ett pedagogiskt verktyg på ett effektivt sätt.

Få studier har publicerats som specifikt fokuserar på användning av AI vid bedömning. I en studie utförd av De Winter et al. (2023, s.4) undersöktes hur AI-modellen GPT-4 kunde användas för att bedöma en 460 ord lång sammanfattning av en studentskriven uppsats. Resultaten visade att med rätt promptar kunde GPT-4 generera användbar feedback och upptäcka både stilistiska och innehållsmässiga fel i texten. Trots detta konstaterades att GPT-4 inte var tillräckligt pålitlig för att sätta betyg på texterna och att mänsklig granskning fortfarande var nödvändig (s.11). Användningen av GPT-4 diskuterades även ur ett etiskt perspektiv med frågor om rättvisa, integritet och påverkan på traditionella utbildningsvärden.

Rudolph et al. (2023, s.352) presenterar i sin forskningsöversikt resultat och slutsatser

om hur ChatGPT kan användas vid bedömning av studenters texter vid universitet och högskola. Rudolph et al. menar att AI är på väg att bli etablerat och att det är viktigt att följa och engagera sig i detta snabbt utvecklande område. Enligt Rudolph et al. är det nödvändigt att lärare anpassar sina undervisnings- och bedömningsmetoder. Diskussioner från undervisningsväsendet har gått från att förbjuda eller förhindra användningen av verktyget till att diskutera om det bör inkluderas i läroplanerna. Det beräknas ta 10 år för skolväsendet att anpassa sig efter denna nya verklighet: två år för studenter att förstå verktyget, ytterligare tre år för lärare att inse att studenter använder verktygen och sedan fem år för universitet att bestämma vad de, om något, ska göra åt användandet av det (Rudolph et al. 2023, s.354).

3 Teoretisk inramning

I detta kapitel presenteras studiens teoretiska inramning, formativ bedömning.

3.1 Formativ bedömning

Formativ bedömning är, enligt Sadler (1989, s.135), en avgörande pedagogisk strategi för att främja elevens lärande och utveckling. Bedömningen bör inte bara vara en mätning av prestation utan snarare en process där feedback används för att guida eleverna. Sadlers teori om formativ bedömning innebär en process där eleverna får återkoppling på sin prestation för att förbättra sitt lärande. Sadler (1989) definierar bedömning som processen och feedbacken som den konkreta stöttningen som hjälper eleverna i processen. Detta skiljer sig från summativ feedback, som fokuserar på att bedöma resultatet av lärandet. I den formativa feedback-processen är syftet att identifiera och revidera brister under lärandets gång för att främja utveckling. När man ger formativ feedback är det viktigt att den är konkret, specifik och kopplad till målen för lärandet. Det innebär att feedbacken bör fokusera på specifika aspekter av elevens prestation och ge tydliga anvisningar om hur eleven kan förbättra sin text.

Enligt Sadler (1989, s.122) bör även feedbacken vara inriktad på den process som ledde till prestationen och inte bara själva resultatet. Det är också viktigt att feedbacken ges i tid, så att eleverna har möjlighet att använda den under lärprocessen. Att ge formativ feedback kan vara en utmaning på grund av den tid och det engagemang som krävs av läraren. Att ge meningsfull feedback förutsätter både kunskap om ämnet och förmåga att kommunicera effektivt med eleverna. Dessutom kan det vara utmanande att ge feedback som är specifik och relevant för varje elevs behov. Trots utmaningar har formativ feedback visat sig ge goda resultat. När läraren ger återkoppling på elevers arbete kan brister i elevens förståelse eller prestationsförmåga identifieras och åtgärdas. Detta kan leda till ökad motivation och engagemang för lärandet, eftersom eleverna ser konkreta möjligheter till förbättring. Dessutom kan formativ feedback främja en kultur där lärandet ses som en process av kontinuerlig förbättring, snarare än av bedömning av prestationer. Formativ bedömning syftar till att skapa en dynamisk och interaktiv undervisningsmiljö där både lärare och elever deltar aktivt och lär sig av bedömningsprocessen (Sadler, 1989, s.141).

Black och Wiliam (1998) har fördjupat förståelsen för formativ bedömning och utvecklat begreppet genom att betona vikten av klassrumsbedömning och elevengagemang i sina studier. De menar att formativ bedömning inte bara handlar om att lärare ger feedback till

eleverna utan vill också att eleverna ska uppmärksamma sina egna framsteg genom självbedömning och sätta sina egna mål för lärandet (s.6). När eleverna aktivt deltar i bedömningsprocessen ökar den formativa bedömningens effekt och främjar en fördjupad förståelse, enligt Black och Wiliam. En annan grundläggande idé är att bedömningen inte ska vara en isolerad händelse utan en kontinuerlig aktivitet där läraren samlar information, tolkar resultat och anpassar undervisningen baserat på detta. Detta kallas också för formativ undervisning (Black & Wiliam, 1998, s.8).

3.1.1 Teorins tillämpning

Sadlers teori skapar en ram för att förstå hur AI-feedbacken kan främja eller hindra elevernas lärande och skrivutveckling, samt identifiera pedagogiska utmaningar och möjligheter som uppstår vid användning av AI-bedömning inom skrivundervisning. Genom att identifiera mönster och trender i feedbacken samt analysera de reviderade texterna kommer slutsatser att dras om hur väl elevernas prestationer matchar de förväntade målen och huruvida feedbacken stödjer eleverna i deras skrivutveckling.

Jakobson (2019, s.13–14) förklarar att inom utbildningsforskning används ordet feedback för att beskriva information som ges till elevers prestationer. Ofta används det tillsammans med ordet *formativ*, och fungerar som en brygga mellan undervisning och elevens skrivutveckling. Den distinktion som Sadler gör mellan formativ bedömning och formativ feedback görs också i denna studie, där formativ bedömning syftar på förhållningssättet som beskriver en process, det vill säga inte själva återkopplingen i sig. Däremot används formativ feedback som synonym till feedback, respons och återkoppling, där formativ feedback alltid syftar till att vara framåtsyftande.

4 Metod och material

Denna studie tillämpar kvalitativa metoder för att utforska effekter av och elevers föreställningar om formativ feedback från artificiell intelligens (AI) inom svenskundervisning för årskurs 1 på gymnasiet. Textanalys används för att dels granska den respons som AI ger, dels undersöka vilka revideringar som elever genomför med hjälp av respons från AI. Vidare har semistrukturerade gruppintervjuer genomförts för att ge inblick i elevernas upplevelser av att få AI-genererad feedback jämfört med traditionell återkoppling. Genom att kombinera textanalys och kvalitativa intervjuer strävar studien efter att ge ökad förståelse av AI-bedömningens användbarhet och mottagande i skolmiljön.

Tabell 1: Materialöversikt

Material	Debattartiklar	AI-genererad feedback	Reviderad text i debattartiklar	Transkriptioner av gruppintervjuer
Antal	24	24	24	5
Antal ord	15 753	7 822	2 190	8 250
Genomsnitt antal ord	656,4	325,9	91,25	1650

4.1 Datainsamling och urval

Datainsamlingen har genomförts i svenskundervisningen i två gymnasieklasser. De 24 deltagande eleverna representerade en blandad grupp av ett yrkesförberedande- och ett högskoleförberedande program.

4.1.1 Insamling av texter

I svenskundervisning för årskurs 1 fick 24 elever i uppgift att skriva debattartiklar. Uppgiften (se Bilaga 1) innebar att eleverna skulle välja ett ämne som intresserade dem och kraven var att formulera en tydlig tes och presentera tre argument som stödde deras ståndpunkt, samt ett motargument som de skulle bemöta. Debattartikeln skulle även innehålla en tydlig inledning och avslutning. Eleverna uppmanades även att använda sig av ett språk som följer

standardnormen för svenskt språkbruk. Inför denna uppgift fick eleverna explicit undervisning om textgenren argumenterande text med fokus på debattartikel. Eleverna fick även en lathund samt två skrivmallar som stöd i skrivprocessen. Eleverna fick 2–3 lektioner (à 60 minuter) för att skriva sina debattartiklar innan de fick respons från AI. När de fick tillbaka sina debattartiklar förväntades eleverna revidera två punkter från den AI-genererade responsen. För revideringen avsattes 1,5 lektion (ca 90 minuter).

I samband med att skrivuppgiften presenterades fick eleverna information om föreliggande studie. De fick veta att deltagande innebar medverkan i gruppintervjuer och användning av deras texter. Samtliga elever valde att delta och blev informerade om sina rättigheter samt förväntningar som knöts till deras deltagande, se vidare 4.3 Etiska överväganden.

4.1.2 AI-bedömning

För att genomföra AI-bedömningen användes ChatGPT 3.5, en digital avancerad textgenererande AI-modell. Det är denna specifika version "ChatGPT" samt "AI" syftar till när de benämns i texten. Detta verktyg är kostnadsfritt och tillgängligt online. ChatGPT har förmåga att förstå komplexa språkstrukturer och kan utifrån sin förmåga bedöma texter mot en fördefinierad prompt (OpenAI, n.d.a). Prompt syftar på den information som ges till AI som inkluderar direktiv om vad den förväntas göra, till exempel generera en text eller svara på en specifik fråga. Detta innebar att AI-modellen bedömde hur väl varje elevtext uppfyllde kriterier utifrån uppgiftsbeskrivningen och gav en objektiv bedömning.

I processen att generera AI-feedback överförde jag varje elevtext individuellt till ChatGPT inklusive en prompt (se Bilaga 1) som specificerade vad AI:n förväntades göra med texten. Målet var att säkerställa en adekvat och likvärdig respons var uppgiftsinstruktionerna samt förväntningarna som gavs till ChatGPT precisa och tydliga. Prompten var influerad av en bedömningsmatris för nationella prov för ämnet svenska (se Bilaga 2). Denna matris fokuserar på fyra övergripande aspekter: *helhetsbedömning, innehåll och textanvändning, disposition och sammanhang* respektive *språk och stil*. Dessa aspekter förtydligades för ChatGPT i olika kvaliteter. *Helhetsbedömning*: anpassning till syfte, mottagare och situation samt självbärande text, *innehåll och textanvändning*: aktiviteter och innehållslig komplexitet, *disposition och sammanhang*: textbindning, inledning, avslutning, styckeindelning, styckemarkering och omfång samt *språk och stil:* skriftspråksnormer, variation, välformulerat språk och situationsanpassning. Nyckelordet i denna process är *formativ feedback*, vilket

innebär att ChatGPT inte bara identifierade textens svagheter och gav konstruktiva rekommendationer för hur dessa kunde åtgärdas utan även lyfte fram textens olika kvaliteter. Denna förväntan förtydligades explicit i prompten. En prompt provades först på en motsvarande elevtext, det vill säga en debattartikel skriven av gymnasieelev för att säkerställa åldersadekvat feedback. Debattartikeln som användes erhölls genom en tidigare anställning som lärare, där den skrevs av en av mina dåvarande elever. Denna prompt reviderades senare två gånger innan den tillämpades på deltagarnas debattartiklar med bland annat förtydligande av de fyra aspekterna.

Responsen på debattartikeln tog bara några sekunder för AI att generera och den formativa feedbacken var tydligt strukturerad. I genomsnitt var längden på den AI-genererade responsen 325,9 ord (för exempel, se Bilaga 3 och 4.). Kommentaren inleddes med helhetsbedömning, en övergripande bedömning av debattartikeln. Där lyftes i första hand styrkor men även en slags diagnos om texten var självbärande eller inte. Därefter lyftes de andra tre aspekterna i varsitt stycke: innehåll och textanvändning, disposition och sammanhang samt språk och stil. Under varje aspekt beskrevs elevens prestation, först styrkor, därefter förbättringsförslag. Kommentaren avslutades med tre exempel på förbättringar. Här konkretiserades hur eleverna kan utveckla texten. Feedbacken som AI genererade kopierades och lades in längst ner i varje elevs debattartikel och delades på så sätt med eleverna. Varje elevtext tilldelades minst tre och högst sex förbättringsförslag. Antalet önskade kommentarer stod inte med i prompten utan var någonting som ChatGPT själv tog beslut om. Eleverna uppmanades att revidera sin text utifrån minst två av dessa förslag. Valet av punkter att revidera överläts till eleverna själva. Eleverna fick instruktionen att färgmarkera ändringar i sina texter för att underlätta granskning och för att ge en överblick över hur mycket och vad varje elev valt att ändra.

En initial bedömning genomfördes av elevtexterna innan analys av revideringar för att möjliggöra en jämförelseanalys om textens kvalitet utvecklats eller inte efter revidering. Detta gjordes genom en bedömningsmatris (se Bilaga 4).

4.1.3 Gruppintervjuer

Efter respons och revidering genomfördes gruppintervjuer med eleverna för att utforska deras upplevelse av att erhålla AI-genererad feedback på sina debattartiklar. Gruppintervjuerna genomfördes i fem grupper om tre till sex elever för att möjliggöra en bredare diskussion och en öppen dialog kring deras upplevelser. Intervjuguiden (se Bilaga 5) utformades för att täcka

olika aspekter av elevernas upplevelse och för att ge inblick i hur de tolkade och använde feedbacken från AI för att förbättra sina texter. Frågorna berörde ämnen som känslor vid mottagande av feedback, jämförelser mellan AI- och lärarfeedback samt användbarhet hos AI-feedbacken, samt vilken typ av feedback som var lätt eller svår att tolka och implementera på egen hand. Eleverna ombads även att reflektera över hur feedbacken från AI påverkade kvaliteten på deras debattartiklar och om det fanns områden där de har önskat mer detaljerad feedback. Ytterligare frågor riktades mot elevernas uppfattning om huruvida AI-feedbacken bidrog till deras skrivutveckling och om det fanns några specifika tankar eller reflektioner de ville dela om sin erfarenhet av att arbeta med AI-genererad feedback. Intervjuerna spelades in via mobiltelefon och transkriberades för att möjliggöra den kvalitativa analysen av elevernas respons och synpunkter.

Intervjuer valdes som en lämplig metod för denna studie av flera skäl. Intervjun möjliggör en djupare förståelse av deltagarnas attityder och upplevelser av ett fenomen (Lagerholm, 2024, s.68–69). Det förklaras bland annat av att intervjuer erbjuder en flexibilitet som tillåter en anpassning till situationen och en möjlighet för informanterna att påverka intervjuprocessen. Detta är särskilt värdefullt när man undersöker personliga upplevelser genom att deltagarna har möjlighet att förtydliga eller understryka saker. De har även möjlighet att reda ut potentiella oklarheter med frågan. Intervjuer möjliggör alltså en mer detaljerad undersökning av elevernas responsarbete, vilket är särskilt relevant för att fånga nyanser i deras upplevelser. De erbjuder också en möjlighet att utforska och förstå elevernas tankar, vilket kan vara svårt att uppnå genom kvantitativa metoder som enkäter.

4.2 Analys av data

I följande avsnitt beskrivs hur dataanalysen genomfördes för att undersöka effekten av AI-genererad feedback på elevers skrivande. Denna analys omfattar en kvalitativ innehållsanalys av den AI-genererade feedbacken och elevernas revideringar av sina texter, samt en tematisk analys av gruppintervjuer. Etiska överväganden som har följts under studien diskuteras också.

4.2.1 Kvalitativ innehållsanalys av AI-genererad feedback och revidering av elevtext

I ett första steg genomförde jag en granskning av feedbacken för varje enskild debattartikel för att identifiera vilken typ av återkoppling som gavs inom respektive aspekt:

helhetsbedömning, innehåll och textanvändning, disposition och sammanhang samt språk och stil. För aspekten helhetsbedömning gav ChatGPT inga förbättringsförslag, medan den formulerade förslag till förbättringar för övriga tre aspekter. Efter en granskning av samtliga förbättringsförslag identifierades femton olika feedback-kategorier (se Tabell 2). Kommentarer som efterfrågade samma typ av revidering kategoriserades tillsammans, exempelvis “undvik upprepningar” och “variera formuleringar”. *Disposition och sammanhang* var den aspekt som hade minst antal förbättringspunkter, endast tre stycken. *Innehåll och textanvändning* inkluderade fem olika förbättringspunkter medan språk och stil hade åtta olika förbättringspunkter.

Tabell 2: Feedback-kategorier inom de tre aspekterna: innehåll och textanvändning, disposition och sammanhang samt språk och stil.

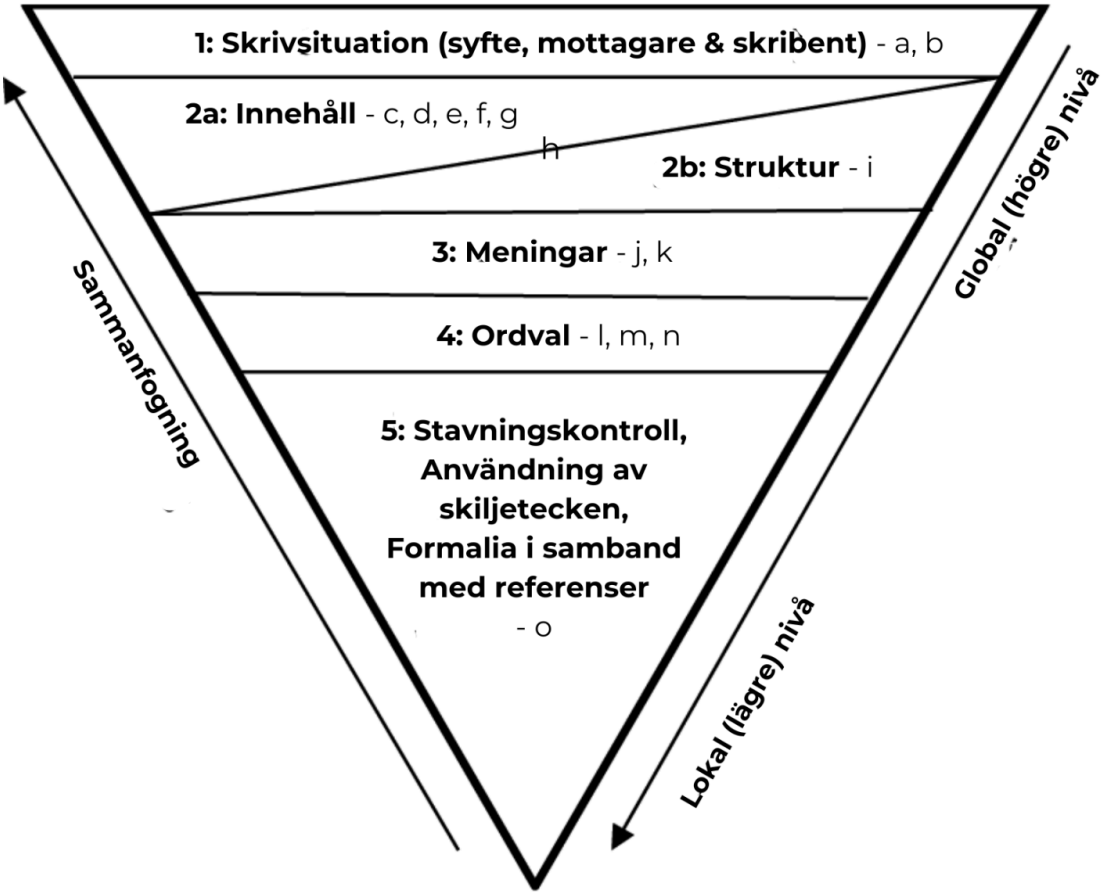
Innehåll och textanvändning	**Disposition och sammanhang**	**Språk och stil**
Inkludera forskning/statistik/expert utlåtanden	Använd mer övergångsord/-fraser eller klargörande meningar	Variera språkliga tekniker och stilistiska grepp
Bemöt motargument/inkludera perspektiv	Dela upp texten i stycken	Anpassa språk och stil
Fördjupa reflektion och analys	Förtydliga inledning och/eller avslutning	Undvik grammatiska fel/språkliga brister
Inkludera mer information/exempel/ytterligare detaljer		Inkludera en mer personlig ton
Förtydliga förklaringar/exempel		Undvik upprepningar/variera formuleringar
		Bryt ner långa meningar
		Använd känsloladdade ord eller retoriska frågor
5	**3**	**7**

Efter en inledande kategorisering användes Dysthes et al. (2011) texttriangel som ett verktyg för att vidare knyta de femton kategorierna till olika textnivåer. De femton olika kategorierna

kopplades då till de fem textnivåerna i texttriangeln: *skrivsituation, innehåll, struktur, meningar, ordval* samt *formalia*. Texttriangeln erbjuder en ram för att kategorisera kommentarerna baserat på deras relativa position längs en skala mellan ett globalt och ett lokalt textfokus. De två översta nivåerna som innefattar skrivsituation, innehåll samt struktur betraktas som globala eftersom de behandlar mer övergripande aspekter av texten. De tre nedersta nivåerna som inbegriper meningar, ordval samt formalia ses som lokala då de fokuserar på specifika detaljer i texten. Även om texttriangeln olika delar kan tolkas som ett kontinuum snarare än som ett antal fasta nivåer, har det varit ändamålsenligt för denna analys att dela in kommentarerna i separata kategorier och nivåer. Detta möjliggjorde en förståelse av hur feedbacken kunde kopplas till antingen en mer lokal eller global nivå i texten. Samtliga textnivåer identifierades bland de AI-genererade kommentarerna och uppdelningen bestod av nio kategorier på global nivå samt sex kategorier på lokal nivå (se Figur 1). I ett andra steg analyserades elevernas revideringar. Under elevernas revideringsprocess uppmanades de att färgmarkera sina ändringar i texten för att ge en snabb och tydlig översikt över dessa ändringar. I samband med detta ställdes kravet att varje elev skulle revidera minst två av de punkter som tilldelats på deras individuella texter.

GLOBAL NIVÅ

a. Använd en mer personlig ton
b. Anpassa språk och stil
c. Fördjupa reflektion och analys
d. Inkludera forskning/statistik/expertutlåtanden
e. Bemöt motargument/inkludera fler perspektiv
f. Förtydliga förklaringar/exempel
g. Inkludera mer information/exempel/ytterligare detaljer
h. Förtydliga inledning och/eller avslutning
i. Dela upp texten i stycken



LOKAL NIVÅ

j. Använd mer övergångsord/-fraser eller klargör meningar
k. Bryt ner långa meningar
l. Använd känsloladdade ord eller retoriska frågor
m. Undvik upprepningar/variera formuleringar
n. Variera språkliga tekniker och stilistiska grepp
o. Undvik grammatiska fel/språkliga brister

Figur 1: Textnivåer. Bearbetning av Dysthes et al. (2011, s.41) texttriangel.

Under nästa steg skapades ett kalkylark för all respons som eleverna fick samt vilken respons eleverna valde att åtgärda. Varje debattartikel numrerades och kategoriseringen i nivåer

användes för att markera vilka feedback-kategorier som gavs på respektive debattartikel. I ett tredje steg analyserades elevernas reviderade texter. Debattartiklarna granskades för att identifiera vilka feedback-kommentarer som tagits i beaktande vid omarbetningen. Ett nytt kalkylark användes för att analysera och dokumentera dessa observationer. I ett fjärde steg kategoriserades de reviderade debattartiklarna i tre kvalitetsnivåer: goda lösningar, medelgoda lösningar och mindre goda lösningar. För att genomföra denna klassificering analyserades först respektive utkast och jämfördes sedan med den reviderade versionen. I denna fas analyserades texterna om dess kvalitet hade utvecklats eller inte. Goda lösningar bedömdes hålla hög kvalitet och hade en hög utveckling på sin text. Medelgoda lösningar visade vissa förbättringar men hade inte utvecklats i lika hög grad som de texter med goda lösningar. Mindre goda lösningar var de som antingen försämrades jämfört med ursprungsversionen eller som fortfarande höll låg kvalitet alltså hade ingen eller marginell textutveckling.

4.2.2 Tematisk analys av gruppintervjuer

Braun och Clarke (2006) har etablerat en metod för tematisk analys som används frekvent inom kvalitativ forskning. Deras metod ger en strukturerad process för att identifiera, analysera och rapportera teman i data. För att genomföra den tematiska analysen enligt Braun och Clarkes metod följdes deras steg-för-steg process. Innan analysen påbörjas är det viktigt att få en övergripande förståelse för datamaterialet, vilket inkluderade att läsa igenom transkriptionerna från gruppintervjuerna. Först lästes därför de fem transkriptionerna noggrant för att skapa en initial förståelse för innehållet och de olika perspektiven som eleverna förde fram. Under denna inledande fas dokumenterades spontana reflektioner och tankar. Dessa reflektioner, som inte följde en strikt metodik, utgjorde en första utforskning av datamaterialet och syftade till att fånga omedelbara observationer inför den fortsatta analysen.

Under fas två, vars fokus är initial kodning, skapades en grundläggande förståelse för datan genom att identifiera och markera segment av texten som verkade relevanta för forskningsfrågorna samt notera intressanta mönster. I denna fas gjordes en uppdelning av positiva och negativa associationer och aspekter med feedbacken.

Under fas tre analyserades de markerade segmenten för att identifiera mönster och gemensamma teman. Först noterades frekvensen av återkommande teman som diskuterades av deltagarna. Detta gav en uppfattning om vilka teman som var mest framträdande och relevanta för gruppen. Vidare observerades liknande upplevelser eller åsikter som delades av flera deltagare. Detta indikerade gemensamma uppfattningar inom gruppen och bidrog till en

mer samlad bild av deras perspektiv. Samtidigt var också uppmärksamhet riktad mot eventuella skillnader och variationer i deltagarnas åsikter. Detta tillät en belysning av olika perspektiv och erfarenheter inom gruppen samt förståelse av nyanserna i deras synpunkter. Från den initiala uppdelningen av positiva och negativa aspekter skapades sex primära teman: förändring, användbarhet, effektivitet, begriplighet, tillförlitlighet och tilltal.

Under fas fyra, som syftar på att granska sina data, definierades varje tema i sin specifika kontext, det vill säga i transkriptionerna. Syftet var att kontrollera att det inte saknades något tema samt ge en översiktlig bild av den tematiska analysen. Två representativa exempel valdes ut från intervjuerna för att representera respektive tema. Detta bidrog till att ge en djupare förståelse för kärnan i respektive tema och dess koppling till deltagarnas erfarenheter och åsikter.

I analysens sista fas gjordes en sammanhängande beskrivning av varje identifierat tema (se Tabell 3). Dessutom gjordes en ny granskning av transkriptionerna för att beräkna hur ofta olika teman förekom i intervjuerna som ett sätt att identifiera de mest relevanta temana. I resultatredovisningen presenteras temana i denna ordning. De tre vanligaste temana var *användbarhet* som förekom 115 gånger, *effektivitet* som noterades 54 gånger och *begriplighet* som observerades 43 gånger.

Tabell 3. Definiering av teman från gruppintervjuer

Tema	**Beskrivning**	**Exempel**
Användbarhet	Under detta tema diskuterar eleverna den praktiska och pedagogiska användbarheten av feedbacken. Det inkluderar bedömningen av hur och om feedbacken har hjälpt eleverna att förbättra sina texter. Framför allt diskuterar eleverna om och i så fall hur feedbacken har hjälpt dem att förbättra sina texter.	Elev 3: Den här hjälpte ju lite, det här ska vara med och det här ska vara med. Men inte riktigt hur man förbättrar texten. Men jag tyckte att den hjälpte lite grann. (Intervju 1) Elev 2: Jag tyckte ändå det var bra. Det är ju kul att se vad man kan göra bättre i stort liksom. (Intervju 2)
Effektivitet	Eleverna diskuterar feedbackens omfattning och kvalitet. Eleverna diskuterar även hur väl AI-responsen uppfyller sitt syfte att stödja deras lärande och skrivutveckling.	Elev 3: Alltså, det var ju inte dåligt. Det var ju ändå, vi fick ändå ganska mycket feedback. (Intervju 4) Elev 3: Jag tyckte också det var bra. Det kom ju väldigt fort och ganska mycket. (Intervju 1)
Begriplighet	I detta tema diskuteras feedbackens tydlighet och användbarhet för eleverna. Fokus ligger på hur lättförståelig, tydlig, konkret och anpassad feedbacken är till elevernas förståelse och kunskapsnivå.	Elev 1: Vissa var ju lite oklara men då valde jag att inte eh, jobba med de förbättringarna utan jag jobbade med några andra som jag förstod och kunde lösa själv liksom. (Intervju 1) Elev 2: Man förstod inte riktigt vad är det den vill att jag ska göra. Och hur. (Intervju 3) Elev 1: Jag förstod väl det som stod där. Alltså att jag fick det på att jag skulle använda mer retoriska frågor och mer övergångsfraser eller vad det var.

		Det var det jag använde och det kändes väl ganska tydligt som jag skulle förbättra. (Intervju 4)
Tillförlitlighet	Diskussionen fokuserar på den upplevda graden av pålitlighet och relevans Eleverna bedömer feedbacken utifrån dess objektivitet, korrekthet och överensstämmelse med förväntningar på relevans och noggrannhet.	Elev 2: Men sen så tyckte jag väl att vissa grejer, att jag inte riktigt höll med om det. (Intervju 2) Elev 2: Och jag fick ju att jag ska (skriva en tydlig inledning som introducerar ämnet och tes på ett engagerat sätt. Och sen så ska jag göra en bättre sammanfattning också. Men jag tyckte att min inledning och avslutning var ganska bra. Och jag slutade med "avslutningsvis" och jag gjorde en, en sista kommentar. (Intervju 2) Elev 3: Ja. Det var ju svårt eftersom att feedbacken var ju något som man redan hade i texten som inte behövde förbättras. (Intervju 3)
Tilltal	Under detta tema utforskar eleverna hur feedbacken presenteras och kommuniceras. Det inkluderar bedömningen av feedbackens ton, stil och personliga relevans.	Elev 3: Jag tycker att det hade varit bättre från en lärare för här fick ju nästan alla precis samma. En lärare kanske är lite mer olika och sånt. (Intervju 1) Elev 3: Nej men, det var samma sak tycker jag. Också det att den skrev samma till alla. Det gjorde att man kände att ah, det här är bara allmänt. Det är inte egentligen vad min text behöver. Och inte så jättefördjupande och detaljerade vad man skulle göra,

		sen var det ju fortfarande liksom... Det fanns ju att det här behöver förbättras och så, men det var inte riktigt vad jag kände kunde hjälpa så mycket. (Intervju 1) Elev 1: När alla fick samma så kändes det som att... Det kanske inte var det man behövde förbättra om alla fick samma, om alla hade samma problem. Det var lite så det kändes. Eller i alla fall, jag och två andra vi hade exakt samma feedback liksom. Så borde det kanske inte vara egentligen. (Intervju 4)
Förändring	Här diskuteras elevernas attityder gentemot införandet av nya teknologier och metoder, som AI-genererad feedback.	Elev 2: Jag tycker inte att man ska använda det i skolan. AI är ju en förändringsgrej. (Intervju 1)

4.3 Etiska överväganden

I denna studie har jag följt de fyra huvudkraven för etisk forskning med människor enligt riktlinjerna från Vetenskapsrådet (2017). Informationskravet tillämpades genom att eleverna fick tydlig och omfattande information om studiens syfte samt möjligheter att ställa frågor. Eleverna fick även ta del av vilka förväntningar som fanns på medverkan för den som valde att delta i studien, nämligen att deras debattartiklar användes för analys samt att de skulle delta i en efterföljande gruppintervju. Eleverna informerades om att deras deltagande var frivilligt och att de hade rätt att avstå från fortsatt medverkan när som helst under studiens gång.

Samtyckeskravet uppfylldes genom att samtliga deltagare fick en fysisk samtyckesblankett att skriva under. Samtliga deltagare är över 15 år och kunde själva ge sitt samtycke till medverkan. På blanketten förtydligades studiens syfte, förväntningar, hantering av kontaktuppgifter samt kontaktinformation om det fanns några vidare frågor eller

funderingar kring deras medverkan.

Det tredje kravet, konfidentialitetskravet, säkerställde att all insamlad data har behandlats konfidentiellt och deltagarnas identiteter har skyddats. Debattartiklarna har kodats som siffror 1–24 och en kodnyckel, som enbart finns i fysisk form och kommer att slängas efter att studien avslutats och rapporterats. Intervjuerna är sparade på universitetets Google Drive och har raderats från privat externa enheter. Transkriberingarna har pseudonymerats genom att deltagarna kodades som siffror 1–6 i varje intervju. Även här finns en fysisk kodnyckel på papper som kommer att slängas. Ingen elektronisk data kommer att sparas när studien är avslutad.

Studien följer även det sista kravet, nyttjandekravet, vilket innebär att data som samlats in endast används för det angivna syftet. Detta säkerställer att deltagarnas bidrag till forskningen respekteras och används på ett etiskt sätt. Denna studie har också följt GDPR:s bestämmelser och säkerställt att all persondata hanterats i enlighet med dataskyddsförordningens krav.

5 Resultat

I detta kapitel presenteras svaren på studiens forskningsfrågor. I avsnitt 5.1 och 5.2 redovisas svaren på de första två frågorna: *På vilka sätt kan AI ge formativ bedömning av elevtexter?* samt *Hur påverkar AI-genererad formativ feedback elevers revidering av skrivna texter?*. Därefter i avsnitt 5.3 redovisas svaren på studiens tredje forskningsfråga: *Hur upplever elever att få AI-genererad feedback?*. Resultatdelen avslutas med avsnitt 5.4, en sammanfattning av studiens mest centrala resultat. Sammanfattningen fungerar också som en övergång till diskussionen, där även studiens sista forskningsfråga behandlas: *Vilka pedagogiska möjligheter och utmaningar kan identifieras vid användning av AI-bedömning inom skrivundervisning?*.

5.1 Övergripande bild av ChatGPTs feedback och elevernas revidering

I tabell 4 ges en översiktlig bild av vilken typ av feedback som AI genererat som ett svar på den prompt som använts samt vilken feedback som eleverna har valt att använda vid sina revideringar. Tabellen följer texttriangelns inledning av nivåer (se Figur 1) och skiljer särskilt mellan global och lokal textnivå. Av tabellen framgår att feedbacken från ChatGPT fördelas relativt jämnt mellan en global nivå (39 förslag på revideringar) och en lokal nivå (46 förslag på revideringar). Detta antyder att eleverna har fått mer detaljerad feedback som rör specifika aspekter av deras texter. Enligt Dysthe et al. (2011, s.187–188) innebär en global nivå övergripande aspekter såsom ämnesfokus, struktur och sammanhang. En mer lokal nivå fokuserar mer på detaljer såsom grammatik och ordval, alltså språklig utformning.

På global nivå uppmanas elever framförallt att “inkludera forskning/statistik/expertutlåtanden” (13 st) medan bara en elev får kommentaren att “anpassa språk och stil” och “dela upp texten i stycken”. Den respons som är vanligast förekommande på lokal nivå är “använd mer övergångsord/-fraser eller klargör meningar” (21 st) medan den minst frekventa är “bryt ner långa meningar” (2 st). Varje debattartikel fick minst tre kommentarer men några fick fler, upp till sex kommentarer.

Eleverna har främst fokuserat på att bearbeta feedback på lokal nivå (32 st). Den typ av feedback som eleverna reviderat mest på lokal nivå är använd mer övergångsord/-fraser eller klargör meningar (15 st). På global nivå görs totalt 16 revideringar. Framför allt har eleverna valt att revidera sina texter genom att inkludera forskning/statistik/expertutlåtanden (7) samt inkludera mer information/exempel/ytterligare detaljer (5). Det fanns också typer av

feedback på global nivå som ingen elev valde att använda sig av för sin revidering: anpassa språk och stil, förtydliga förklaringar/exempel samt förtydliga inledning och/eller avslutning.

Tabell 4: Fördelning av AI-genererade kommentarer efter antal elever, revideringar och textnivå enligt texttriangeln (Dysthe et al. 2011).

Respons	**Antal elever som fått respons**	**Antal elever som reviderat aktuell respons**	**Textnivå i texttriangeln**
Använd en mer personlig ton	4	0	Global:1
Anpassa språk och stil	1	0	Global:1
Fördjupa reflektion och analys	3	2	Global:2a
Inkludera forskning/statistik/expertutlåtanden	13	7	Global:2a
Bemöt motargument/inkludera perspektiv	6	1	Global:2a
Förtydliga förklaringar/exempel	2	0	Global:2a
Inkludera mer information/exempel/ytterligare detaljer	6	5	Global:2a
Förtydliga inledning och/eller avslutning	3	0	Global:2a/b
Dela upp texten i stycken	1	1	Global:2b
Summa	39	16	9
Använd mer övergångsord/-fraser eller klargör meningar	21	15	Lokal:3
Bryt ner långa meningar	2	2	Lokal:3

Använd känsloladdade ord eller retoriska frågor	5	5	Lokal:4
Undvika upprepningar/variera formuleringar	6	2	Lokal:4
Variera språkliga tekniker och stilistiska grepp	9	6	Lokal:4
Undvik grammatiska fel/språkliga brister	3	2	Lokal:5
Summa	46	32	6

5.2 Fördjupad bild av elevers revidering

Kravet i undervisningsupplägget var att eleverna skulle revidera minst två punkter av de förbättringsförslag AI har genererat. Eleverna behövde stöd för att tolka vissa kommentarer och vissa uttryck behövde även förtydligas och definieras. Dessutom var det ibland nödvändigt att komplettera feedbacken från ChatGPT med exempel, då eleverna inte alltid förstod kommentarerna. Den initiala tanken var att eleverna skulle jobba självständigt för att undersöka vad de kunde göra med feedbacken utan lärarhjälp, därför genomfördes ingen explicit genomgång av feedbacken i helklass. Vid den inledande fasen av revideringsprocessen uppstod också flera frågor kring vilka kommentarer som skulle beaktas. Detta berodde på att feedbacken från ChatGPT först presenterades utförligt för varje aspekt, exempelvis “Språk och Stil”, följt av tre utvalda punkter som specificerades ytterligare. Eleverna trodde initialt att endast de tre sista punkterna var relevanta för revidering, vilket ledde till osäkerhet kring vilka förbättringar som behövde tas hänsyn till. För att underlätta för eleverna hade jag i förväg markerat samtliga förbättringsförslag i deras respons med understrykningar. Trots detta blev det missförstånd om vad som förväntades. I exempel 1 visas en kommentar från ChatGPT där eleven föreslås att inkludera forskning/statistik/expertutlåtanden i sin text.

> (1) Ditt ämne är relevant och du använder statistik och forskningsresultat för att stödja dina påståenden, vilket är bra. Dock finns det vissa områden där innehållet kan utvecklas för att övertyga läsaren mer effektivt. Exempelvis, när du diskuterar samhällseffekterna av dålig ekonomi, kan du inkludera fler

konkreta exempel eller följderna av ekonomisk osäkerhet i olika samhällssektorer.

Majoriteten av eleverna, totalt 19 stycken, valde att revidera sina debattartiklar enligt det förutbestämda kravet, alltså att använda sig av två punkter från den AI-genererade feedbacken. Två elever valde att endast revidera en punkt medan tre elever valde att använda sig av tre olika punkter från responsen vid sin revidering. Efter att ha systematiskt gått igenom varje debattartikel visade det sig att elevernas revideringsmönster varierade. Två huvudmönster identifierades. Det första mönstret var att revidera texten i den ordning som ChatGPT föreslog, detta gjorde nio elever. Det andra mönstret var att strategiskt välja vilka förbättringspunkter som skulle ändras i sin text, detta gjorde femton elever. Vad som styr dessa val ger intervjuerna vissa svar på, det tycks att eleverna väljer att revidera de punkter som är konkreta, tydliga och lätta att förstå. Vidare visade analysen av elevernas revideringar att omfattningen på dessa varierade avsevärt, med en spridning från 10 ord till 265 ord. Det är dock värt att notera att eleven som reviderade 265 ord ursprungligen hade en text på endast 251 ord. Med tanke på att uppgiftens krav var att skriva en debattartikel på minst 500 ord kan detta påverka tolkningen av revideringsmönstret. I genomsnitt reviderade eleverna sina texter med 91,25 ord, vilket ger en indikation om den relativa omfattningen av deras revideringar.

Utifrån elevernas revideringar delades texterna in i tre nivåer baserat på deras kvalitet och innehåll: goda lösningar, medelgoda lösningar och mindre goda lösningar. Utifrån dessa kategorier har jag beräknat antalet ord som respektive elev reviderat i sin text. Av de 24 eleverna uppdelades 10 i goda lösningar, 7 medelgoda lösningar och 7 mindre goda lösningar. Nivån goda lösningar hade ett genomsnitt på 118,1 antal ord, medelgoda hade ett genomsnitt på 71,4 ord medan de mindre goda lösningarna hade 72,7 ord. En debattartikel i den senare kategorin uppnådde inte kravet på 500 antal ord före revidering. Detta påverkade genomsnittet eftersom eleven lade till 265 ord för att uppnå kravet. Om denna text utesluts från beräkningen, blir det nya genomsnittet för antalet reviderade ord 40,7. Elevernas revideringar visar tydliga skillnader i valet av att revidera globala och lokala förbättringspunkter utifrån deras skrivkompetens. Mer avancerade skribenter (goda lösningar) har reviderat 0,8 globala och 1,4 lokala punkter i genomsnitt. Elever på medelnivå (medelgoda lösningar) har i genomsnitt fokuserat på färre globala aspekter, 0,6, medan de reviderat 1,3 lokala punkter. Den mer grundläggande nivån (mindre goda lösningar) har i genomsnitt endast reviderat 0,4 globala punkter och 1 lokal punkt.

Tabell 5: Kategorisering av kvalitet på lösningar av reviderade elevtexter samt antal globala och lokala revideringar.

Kategori	Goda lösningar	Medelgoda lösningar	Mindre goda lösningar
Antal texter	10	7	7
Antal ord reviderade i genomsnitt	118,1	71,4	72,7 (40,7)
Antal reviderade punkter	8 globala, 14 lokala	4 globala, 9 lokala	3 globala, 9 lokala
Antal reviderade punkter genomsnitt	0.8 global, 1,4 lokal	0,6 global, 1, 3 lokal	0.4 global, 1 lokal

5.2.1 Revideringar på global nivå

I följande två avsnitt kommer varje feedback-kategori från AI att redovisas med exempel på revideringar från deltagarnas debattartiklar. Varje exempel har försetts med benämningar från texttriangeln i form av en bokstav inom parentes. Nedan redovisas de kategorier som elever valt att använda vid revidering, här ingår alltså inte "använd en mer personlig ton", "anpassa språk och stil", "förtydliga förklaringar/exempel" eller "förtydliga inledning och/eller avslutning". I de fall det finns flera exempel, har det valts ut exempel som är representativa för kategorin. Varje kommentar kommer att indikeras med en parentes med bokstav från texttriangeln för tydlighetens skull. Dessa exempel illustrerar hur eleverna har reviderat sina texter för att möta den givna feedbacken. För att ge ytterligare kontext kommer varje exempel att föregås av en kort beskrivning av den tes som debattartikeln argumenterar för.

Under global nivå var det fyra kommentarer från AI som inte användes vid deltagarnas revidering. Dessa var personlig ton (a.), anpassa språk och stil (b.), förtydliga inledning och/eller avslutning (h.) samt förtydliga förklaringar/exempel (f.).

En elev som argumenterade för att estetiska ämnen borde vara obligatoriska på gymnasiet valde att revidera utifrån feedbackpunkten att fördjupa reflektion och analys (c.). Eleven valde att utveckla sina tankar genom att inkludera ett resonemang hur dessa aktiviteter kan bidra till elevers välbefinnande:

> Föreställ dig en gymnasieskola där eleverna inte bara arbetar med matte och historia, utan också är med i konstprojekt och gör fantastiska teateruppträdanden. Dessa olika aktiviteter skapar en mer kreativ miljö och gör att eleverna mår bättre.

En annan elev som valde att använda feedbacken där hen skulle inkludera forskning/statistik/expertutlåtanden i sin text, argumenterade för att minska skärmtiden. I sin revidering inkluderades en referens till en studie av KTH om att blått ljus bidrar till sömnproblem:

> En studie under ledning av en forskargrupp på Harvard har visat att blått ljus dämpar produktionen av det viktiga sömngivande hormonet melatonin. Mobilen är anledningen till att antalet sömnproblem i Sverige bland unga har ökat med 500 procent sedan 2007 till 2019, säger en studie från KTH.

En annan kommentar som en av eleverna valde att revidera utifrån var att bemöt motargument/inkludera perspektiv (e.). Eleven, som argumenterade för att begränsa skärmtiden för barn under 4 år inkluderade en referens till en studie där hen bland annat lyfter att digital medieanvändning främjar mognad inom områden som visuell bearbetning men möter detta perspektiv med att belysa att det samtidigt underutvecklar andra komplexa förmågor:

> I en studie där MRT-skanningar utfördes på 52 friska barn i åldrarna 3 till 5 år, jämfördes hjärnstrukturen efter varje barns digitala medieanvändning. Resultaten visade på ökad mognad inom områden som visuell bearbetning men avslöjade också underutveckling i områden som stödjer mer komplexa färdigheter såsom språkutveckling, läsförmåga och sociala färdigheter.

Att inkludera mer information/exempel/ytterligare detaljer (g.) var en annan punkt från AI som en av eleverna valde att använda vid sin revidering i sin debattartikel om att införa gratis hundbajspåsar. Den AI-genererade feedbacken uppmuntrade eleven till att inkludera ytterligare information om hur hundpajspåsar har påverkat andra kommuner eller städer:

> Filipstads invånare är bosatta i en av de kommuner som har valt att prova på detta, och har märkt en stor skillnad i både miljö- och trivnadsfaktor. Femton stycken “Tikspac” stationer är utplacerade runt om i kommunen, och hjälper ständigt till att sträva efter en trevligare levnadsmiljö. Påsarna är 100% biologiskt nedbrytbara, och är idag utplacerade i hela 370 kommuner, enligt tikspac.com.

Enbart en av deltagarna fick förslaget från AI att dela upp sin text i stycken (i.) med underrubriker. Eleven som argumenterade om att införa obligatorisk undervisning i ekonomi på gymnasiet valde att lägga till flera underrubriker i sin text:

> Din ekonomi påverkar ditt mående (underrubrik tillagd)

5.2.2 Revideringar på lokal nivå

För revideringar på lokal nivå finns exempel på samtliga feedback-kategorier. En av de mest förekommande feedbackkategorierna var att använda mer övergångsord/-fraser eller klargöra meningar (j.) i sin text. En elev som valde att revidera sin text utifrån denna punkt argumenterade för att veganer inte ska överföra sina matvanor på sina husdjur. Eleven valde att dela upp ett stycke i två samt avsluta stycket med att ställa en retorisk fråga för att sedan inleda nästa stycke med att besvara frågan:

> För att det som är menat att gynna djuren inte ska bli upphov till lidande är det viktigt att kunna ta hjälp av professionella som vet vad de pratar om, men vad ska man göra när inte ens proffsen vill ta ställning?
> (Nytt stycke)
> Veterinärer ger inga raka svar. De menar på att det inte finns tillräckligt med bevis för att kunna säga för- eller emot med säkerheten som behövs av en veterinär.

Bryt ner långa meningar (k.) var en annan feedback-kommentar som förekom vid elevers revideringar. I följande exempel där en elev argumenterar för att abort bör vara en mänsklig rättighet delades flera långa meningar upp i två. Revideringen förbättrar inte nödvändigtvis kvaliteten på texten.

> Detta har resulterat i en siffra på 68 000 dödsfall och enligt FN är abort den största mortalitetsorsaken bland gravida kvinnor.
> →
> Detta har resulterat i en siffra på 68 000 dödsfall. Enligt FN är abort den största mortalitetsorsaken bland gravida kvinnor.

En annan elev valde att revidera sin text utifrån kommentaren att använda fler känsloladdade ord eller retoriska frågor (l.). Eleven argumenterade för att stoppa utvecklingen av vissa typer av AI och valde att lägga till ett stycke med en avslutande retorisk fråga vid sin revidering:

> Ett tydligt exempel till detta är användningen av autonoma vapen, där AI-baserade system kan ta beslut om liv och död utan någon som helst mänsklig inblandning, och är det verkligen något som AI ska få ta beslut inom?

På lokal nivå fanns även en kommentar där eleven uppmanades att undvika upprepningar samt variera sina formuleringar (m.). Eleven som argumenterade för att införa obligatorisk

undervisning i ekonomi på gymnasiet (utgjorde ett exempel för att lägga till underrubriker i sin text, se ovan) valde att undvika upprepningar genom att byta ut ordet *kunskap* i flera av sina meningar mot synonymer då det var ett ord som förekom ofta i hens text. Som ett resultat kan denna revidering potentiellt försvaga förståelsen av meningarna då exempelvis "lärdom" inte nödvändigtvis är synonymt med ordet "kunskap" i denna kontext.

> Utan denna *lärdom* blir detta en ond cirkel.
>
> I så fall är det inte bara barn från familjer med bra förutsättningar som får *visdom* om ämnet.

Använda varierande språkliga tekniker och stilistiska grepp (n.) var en kommentar som en av eleverna valde att använda vid sin revidering. Eleven som argumenterade för att minska konsumtionen av tomma kalorier valde att inkludera en liknelse i sin text:

> Det kan liknas med att långsamt och billigt förgifta dig själv eftersom giftet är billigare och smakar godare än motgiftet.

Den sista kommentaren på lokal nivå som användes vid deltagarnas revidering var undvik grammatiska fel och språkliga brister. En elev som skrev om att skolmatens smak och näringsinnehåll borde förbättras gjorde bland andra dessa ändringar i sin text:

> Men det är inte allt i skolan behöver man ha engei* för att orka med skoldagen [...]
> → *energi
>
> koncentrera sig och kunna komma ihåg informationen*
> → *. (punkt har lagts till)
>
> [...]. s*kolmaten
> → *Skolmaten (ändrat till versal)

5.2.3 Sammanfattning av globala och lokala revideringar

Elevernas revideringar på global nivå visar en generell förbättring av deras debattartiklar. De globala revideringarna var ofta fylliga, utförliga och utvecklade texten. Det finns dock undantag, till exempel eleven som fick respons att lägga till underrubriker. Dessa underrubriker tillförde inte några förbättringar till textens kvalitet. På den lokala nivån visas textexemplen att mindre förändringar görs som inte påverkar textens meningsnivå i stor

utsträckning med undantag när elever valde att inkludera retoriska frågor och känsloladdade ord. Ibland gavs förslag som ledde till försämring av texten. En elev uppmanades att undvika upprepningar samt variera sina formuleringar. Eleven försökte följa detta råd genom att byta ut ordet “kunskap” mot synonymer i flera meningar. Denna revidering kan dock potentiellt försvaga förståelsen av meningarna eftersom synonymer som “lärdom” inte nödvändigtvis ersätter “kunskap” i denna specifika kontext.

5.3 Elevernas upplevelse av AI-genererad formativ feedback

Under detta avsnitt besvaras frågeställning 3 genom att presentera de sex teman som framkommit genom den tematiska analysen av gruppintervjuerna. Dessa teman ger en förståelse för olika aspekter av feedbacksprocessen ur elevperspektiv och synliggör deras upplevelser och bedömningar av feedbackens användbarhet och effektivitet. Flertalet elever uttryckte en övervägande positiv inställning när de diskuterade hur det kändes att få AI-genererad feedback. Deltagarna beskrev själva konceptet som roligt och att det kändes bra. Det fanns också elever som uttryckte en neutral attityd.

> Elev 1: Jag tyckte det var roligt. (Intervju 1)
> Elev 2: Nej, det kändes inte så speciellt. (Intervju 4)

5.3.1 Användbarhet

När eleverna diskuterar feedbackens användbarhet ligger fokus på den praktiska användbarheten. Eleverna diskuterar hur feedbacken påverkade deras skrivutveckling och textens kvalitet och innehåll. Under intervjun framkom exempel där feedbacken hjälpte eleverna att identifiera specifika områden för förbättring som de implementerade för att förbättra sina texter. Exempelvis framhöll en deltagare att hen fick återkoppling om att korta ner långa och komplicerade meningar. Eleven gav uttryck för en medvetenhet om behovet av förbättring.

> Elev 3: Just det här med, jag fick en respons att jag skulle variera språket lite mer.
> I: Mm.
> Elev 3: Och det kunde jag ju förbättra lite. Till en viss gräns i alla fall.
> Elev 5: Jag hade samma med övergångar. Alltså då läste man ju hur styckena och det satt ihop. Eller inte. Och så kunde man ju hitta vart det lät lite sämre.

> Elev 2: Och jag hade… Fick att jag hade långa komplicerade meningar. Och då kunde jag se att det kanske blev lite långt vid vissa gånger. Att jag kunde korta ner det liksom. (Intervju 2)

Flera elever gav uttryck för att feedbacken fungerade som en påminnelse om vissa aspekter av debattartikeln, som de hade glömt att inkludera i sitt arbete. Detta indikerar en potentiell styrka hos feedbacken som ett verktyg för att påminna eleverna om viktiga element i uppgiftsinstruktionen. Å andra sidan uttrycker en deltagare också att den allmänna karaktären av feedbacken uppfattas som vag och generell, vilket antyder ett behov av mer konkret och specifik återkoppling. Följande exempel illustrerar detta:

> Elev 4: Men sen tycker jag ändå att den passar bra som en påminnelse. Så att det mesta var ju väldigt generellt feedbacksmässigt. Men så att typ det jag tänker mest på var att förstärka övergångarna. Jag hade glömt att skriva övergångar. Så att det fungerade ju som en bra påminnelse.
> I: Mm.
> Elev 4: På så sätt var det ju bra men jag tyckte att det var väldigt generella grejer som var lite flummiga. (Intervju 2)

Vissa elever gav uttryck för att det var utmanande att använda den AI-genererade feedbacken. Trots att feedbacken erbjöd riktlinjer för förbättringar upplevde flera deltagare svårigheter med att konkretisera och tillämpa dessa direkt på sina texter. Eleverna menade att feedbacken var övergripande och svår att omsätta i praktiken, vilket tyder på en begränsning av dess nytta.

> Elev 1: Jag vet inte, jag tyckte det var svårt att gå in och göra texten bättre när man fick liksom så övergripande [feedback]. Alltså då blir det såhär, då ska man ju inte ändra hela texten utan man ska ju bara ändra lite grann och då vet man ju inte riktigt vad man ska ändra med så övergripande… (Intervju 4)

5.3.2 Effektivitet

Temat effektivitet kretsade kringt den generella omfattningen och hastigheten hos den AI-genererade feedbacken jämfört med traditionell lärarfeedback. Eleverna gav uttryck för att feedbacken var omfattande och täckte många olika aspekter, vilket betraktades som positivt. Då den AI-genererade feedbacken levererades snabbt till eleverna och innehöll omfattande feedback inom de fyra aspekterna: helhetsbedömning, innehåll, disposition och språk, upplevde många deltagare feedbacken som effektiv. Till skillnad från lärarfeedback, som ofta

tar längre tid att få, tog AI-feedbacken bara några sekunder att producera och eleverna fick tillbaka feedbacken på sina texter dagen efter inlämning. Samtidigt har det framkommit att feedbacken emellanåt avvek från de angivna instruktionerna, vilket väckte vissa bekymmer bland deltagarna. Även om elevtexterna redan uppfyllde kravet att möta ett motargument så uppmanade ibland feedbacken till att inkludera ytterligare motargument.

> Elev 2: Jag tyckte den var bra, den tog ju med mycket. Den hade ju innehåll och så hade den disposition och den hade språket. Så det var ju bra att det inte var bara en grej, att det var ganska stort. (Intervju 2)

> Elev 3: Jag tyckte också det var bra. Det kom ju väldigt fort och ganska mycket. Men sen var det också lite annorlunda från instruktionerna du gav till exempel de ville ha mer argument eller andra saker som inte var våra instruktioner liksom. (Intervju 1)

Några elever uttryckte att den AI-genererade feedbacken var bristfällig och önskade att den hade inkluderat konkreta exempel för att förtydliga och illustrera de föreslagna förändringarna. Till exempel påpekade flera deltagare att feedbacken kunde gett explicita exempel på det som den föreslår ska förbättras. Deltagarna uttryckte att detta hade bidragit till att förstå och tillämpa feedbacken mer effektivt. Dessutom framkom det att eleverna saknade grammatisk detaljrättning i feedbacken.

> Elev 1: Den hade ju kanske kunnat ge exempel på, alltså i och med att den ändå läser texten och fattar vad man skriver om så kanske den hade kunnat ge något, alltså konkret exempel på ett motargument. Om jag var osäker på mitt och om det var tillräckligt bra, så hade den kunnat ge exempel på motargument, exempel på argument och typ något medkänsligt, något som inte är sakligt. Alltså typ, ja men som att man kan skriva om något som är blir mer känslosamt. Exempel på det liksom.

> Elev 4: Och som du säger exempel och även… Nu kanske det inte riktigt har med det här att göra men den här rättade ju inte grammatiska grejer eller det som jag sa innan. Så att det tycker jag är bra när lärare gör. [...] (Intervju 3)

5.3.3 Begriplighet

Under temat begriplighet diskuterades elevernas upplevelser av hur lättförståelig feedbacken var för dem. Vissa elever uttryckte att de hade en god förståelse för feedbacken, särskilt när den var konkret och inriktad på specifika förbättringsområden. Ett område som eleverna ansåg vara tydligt och lätt att implementera var uppmaningar att inkludera mer forskning i sin

text.

> Elev 1: Nej jag förstod ganska bra faktiskt. Men det är väl också ganska konkret med forskning. (Intervju 1)

> Elev 1: Jag förstod väl det som stod där. Alltså att jag fick det på att jag skulle använda mer retoriska frågor och mer övergångsfraser eller vad det var. Det var det jag använde och det kändes väl ganska tydligt som jag skulle förbättra.
> Elev 2: Ja samma, jag fick också övergångsfraser och ändra språk lite. Så det gjorde jag. (Intervju 4)

Å ena sidan varierade begripligheten av feedbacken ganska mycket beroende på innehållet i den. Medan vissa förbättringsförslag upplevdes som tydliga och konkreta, såsom att använda fler retoriska frågor eller övergångsfraser, ansåg flera deltagare att det var svårare att tolka och implementera mer generella instruktioner, till exempel att förstärka vissa delar av texten. Dessa deltagare uttryckte viss frustration över den bristande tydligheten samt avsaknaden av konkreta exempel.

> Elev 1: Det var svårt att veta liksom... För samtidigt, du fick... Man fick ju ganska mycket... Eller jag fick det... Feedback, för vad jag kunde förbättra inom vissa grejer. Men det var fortfarande ganska svårt för det var ju ändå jag som liksom skulle veta vad jag behövde liksom ändra inom det.
> I: Mm.
> Elev 1: När det var "förstärk detta". Men hur gör jag det? Jag tyckte det var lite svårt.
> I: Mm.
> Elev 2: Och det tyckte jag med. Hur ska jag göra det här för att ändra det så? För det blev liksom inte att "där" var det du gjorde fel. Så då får man liksom leta i hela texten vad är det som har blivit fel liksom. (Intervju 2)

Under diskussionen kring begriplighet framkom det att vissa deltagare stötte på svårigheter med att förstå viss terminologi som användes. Två exempel som lyftes fram var "skriftspråksnormerna" samt "kommunikativ kvalitet". Denna upplevelse att stöta på ord eller begrepp som eleverna inte ansåg vara adekvata för deras nivå tycks ha skapat förvirring och osäkerhet för eleverna.

> I: Inga konstiga ord eller någonting ni inte visste hur ni skulle applicera på er text? Inga konstigheter?
> Flera: Nej.
> Elev 5: Konstiga ord.

I: Lite konstiga ord… Har du något exempel på…?
Elev 5: Skriftspråksnormerna.
I: Mm.
Elev 5: Det var konstigt.
I: Ja. Är det någon som har något mer exempel som ni kände att det var... Liksom kinesiska nästan.
Elev 4: Kommunikativ kvalitet. (Intervju 5)

5.3.4 Tillförlitlighet

När eleverna diskuterar feedbackens tillförlitlighet framkom olika åsikter bland deltagarna. En del såg positivt på att AI-feedbacken var opartisk, då den bedömer alla elever på samma sätt och inte favoriserar någon specifik elev, vilket eleverna gav uttryck för kan vara en risk när en lärare ger feedback. Å ena sidan påpekade eleverna att en lärare, genom att känna till elevernas historik och tidigare utmaningar, kan ge mer relevant och anpassad feedback än vad en AI kan göra. Dessutom framhölls att AI:n kanske inte kan förstå eller uppskatta elevernas individuella skrivstilar eller stilistiska val på samma sätt som en mänsklig lärare kan göra.

Elev 3: AI är ju så här opartisk om man säger så. Den ser ju inte eleven, läraren kan ju skriva utifrån vilken elev den gillar och så. (Intervju 1)

Elev 3: Han [läraren] vet ju redan hur texterna sett ut tidigare liksom, men det vet ju inte ChatGPT om.
I: Nej, precis.
Elev 2: Alla har ju olika sätt att skriva på också. Jag tänker det är ju lite svårt för ChatGPT att läsa också.
I: Mm
Elev 2: Så vissa saker som man kanske valde att göra, så blev det nej det här tyckte jag inte om. (Intervju 2)

I diskussionen om feedbackens tillförlitlighet framkom både positiva och negativa synpunkter bland deltagarna. En del var kritiska till att feedbacken verkade leta efter fel, även i texter som redan höll en hög kvalitet. De ansåg att vissa förbättringspunkter verkade vara påhittade eller mindre relevanta för textens helhet. Viss feedback stämde inte heller överens med elevernas personliga preferenser om hur texten skulle förbättras. De ansåg att vissa feedbackpunkter var obefogade eller inte korrekta enligt deras bedömning. Å andra sidan fanns det deltagare som höll med om att feedbacken var befogad och var nödvändig för att förbättra texten.

Elev 1: Det är som att den bara liksom bara för att, den måste hitta något som är fel så bara skriver den

något. (Intervju 3)

Elev 2: Jag tyckte inte att det stämde så mycket överens om hur det egentligen skulle göras. Att göra den [texten] bättre. (Intervju 1)

Elev 3: Det stod såhär att jag skulle använda mig i övergångsfraser och det märkte jag ju själv att i styckena att de inte är så följsamma. (Intervju 1)

Samtidigt framkom en kritisk synpunkt från en av deltagarna angående användningen av AI-bedömning, särskilt på personliga texter såsom debattartikel. Deltagaren ifrågasatte tilliten till AI-teknologin, särskilt i sammanhang där det krävs en slags mänsklig känslighet för att både skriva genren samt förstå och bedöma det personliga innehållet.

Elev 6: Ja, jag tänker att det är en väldigt ny teknologi det där. Jag tror inte det ska litas på, speciellt inte på en debattartikel där det inte är som en argumenterande text. För det ska vara lite... Mänskligt, om man säger så och AI är motsatsen. (Intervju 2)

5.3.5 Tilltal

Under temat "tilltal" samlas synpunkter om att eleverna saknar en personlig röst i den AI-genererade feedbacken. Många elever noterade att de fick liknande feedbackpunkter, vilket ledde till en uppfattning om att feedbacken var opersonlig. Flera elever uttryckte att de föredrar feedback från lärare eftersom den förväntas vara mer anpassad till individuella behov och mer varierad.

Elev 3: Jag tycker att det hade varit bättre från en lärare för här fick ju nästan alla precis samma. En lärare kanske är lite mer olika och sånt. (Intervju 1)

Elev 2: Det stod i nästan alla att man skulle ha bättre övergångar tror jag. (Intervju 1)

Elev 1: [...] jag gillar ju lärarens feedback ganska mycket. Det blir ju lite mer personligt. Det här kändes som att alla fick typ samma.
Elev 2: Ja, jag håller med. Det kändes verkligen som att det var opersonligt.
Elev 1: Men läraren då, han läser ju verkligen igenom den och så får man liksom personlig feedback istället för att "ha en övergångsfras". (Intervju 3)

5.3.6 Förändring

Temat "förändring" är ett mycket litet tema. Deltagarna uttryckte att upplevelsen att få AI-genererad feedback var annorlunda och spännande jämfört med traditionell feedback från lärare eller kompisar. Deltagarna beskrev det som nytt och intressant att använda ett digitalt verktyg för att få bedömning av sina texter, vilket var en ny erfarenhet för samtliga elever i studien.

> Elev 1: Eh, nej men jag tyckte det var kul för att det var annorlunda för i alla år har man liksom fått lärare- eller kompisfeedback, så att få liksom digitalt verktyg var... Ja, men det var lite nytt och lite spännande och det kanske är så här det kommer fortsätta. Så det var ju kul att prova på i alla fall. (Intervju 1)

> Elev 1: Eh, alltså… Jag tyckte det var roligt. Jag tyckte att det var… Det är väldigt annorlunda att få öh, feedback från AI då och inte en lärare eller kompis eller så. (Intervju 3)

5.4 Sammanfattning

I detta avsnitt sammanfattas de viktigaste svaren på forskningsfrågorna 1–3. I kommande avsnitt diskuteras dessa resultat samt vilka pedagogiska utmaningar och möjligheter som kan identifieras vid användning av AI-bedömning inom skrivundervisning.

Studien visar att AI kan ge omfattande och snabb formativ feedback på både globala och lokala nivåer av elevtexter. Feedbacken behandlade olika aspekter såsom helhetsbedömning, innehåll och textanvändning, disposition och sammanhang samt språk och stil. Eleverna kunde använda feedbacken för att utveckla sina texter och den genomsnittliga revideringslängden var 91,25 ord. Elevernas revideringsmönster visade en övervägande respons på lokal nivå (46), där de fokuserade på att förbättra specifika detaljer i sina texter framför den globala nivån som fokuserar på övergripande aspekter (39), som ofta är kopplade till textens innehållsliga nivå. Elevernas reviderade artiklar klassificerades i tre nivåer: goda lösningar, medelgoda lösningar och mindre goda lösningar. Genomsnittet av antalet reviderade ord var 118,1 för goda lösningar, 71,4 för medelgoda lösningar och 72,7 (40,7 efter exkludering av en text som inte uppfyllde ord-kravet) för mindre goda lösningar. Eleverna på goda lösningar-nivån har i genomsnitt reviderat 0,8 globala och 1,4 punkter i genomsnitt. För medelgoda lösningar var genomsnittet 0,6 globala och 1,3 lokala punkter. För de elever som hade mindre goda lösningar var genomsnittet för global 0,4 och för lokal 1.

Elevernas upplevelser av AI-feedback var blandade. Vissa fann konceptet spännande

medan andra ansåg det vara opersonligt och ibland svårt att tolka eller använda. Många elever efterfrågade mer personlig och anpassad feedback som endast en mänsklig lärare kan erbjuda.

Vidare öppnar användningen av AI för formativ bedömning inom skrivundervisning upp för flera pedagogiska möjligheter. Den producerar snabb och omfattande återkoppling till elever vilket kan frigöra lärarresurser för andra uppgifter. Samtidigt uppstår utmaningar med att säkerställa att AI-feedbacken är relevant, begriplig och användbar för eleverna. Det krävs också ytterligare forskning och utveckling för att förbättra AI-teknologins förmåga att ge mer personlig och anpassad feedback som kan uppfattas som meningsfull och effektiv för elevernas lärande. Pedagogiska utmaningar och möjligheter diskuteras vidare i nästa kapitel.

6 Diskussion

I detta kapitel kommer de viktigaste resultaten och slutsatserna från studien att diskuteras i relation till tidigare forskning samt didaktiska implikationer som svar på studiens fjärde forskningsfråga *Vilka pedagogiska möjligheter och utmaningar kan identifieras vid användning av AI-bedömning inom skrivundervisning?*. Diskussionen avslutas med en metoddiskussion och några tankar om fortsatt forskning.

6.1 Studiens resultat i relation till tidigare forskning

Studien visar att elever som skriver mer avancerade texter i högre grad fokuserar på globala revideringar, medan elever som skriver på mer grundläggande nivå tenderar att ägna mer tid åt lokala revideringar. Detta är någonting som även konstaterats i tidigare studier, Faigley och Witte (1981) visar till exempel att erfarna skribenter, jämfört med oerfarna, ägnar mer tid att revidera på de globala nivåerna för att förbättra textens struktur och innehåll. Kronholm-Cederberg (2009) noterade också att goda skribenter prioriterar innehållsfrågor och visar en medveten hantering av språket, medan oerfarna skribenter fokuserar på exempelvis meningsbyggnad och ordval på lokal nivå. Resultaten visade även att genomsnittet på antal ord vid revidering skiljde mellan de olika nivåerna. Avancerade skribenter tenderade att göra mer omfattande revideringar. Detta kan relateras till att skickliga skribenter ofta gör makrostrukturella förändringar för att förbättra tydligheten i sin text, vilket kräver en fördjupad förståelse och ofta en större insats i revideringen (Faigley & Witte, 1981).

Deltagarna i denna studie hade svårigheter att använda den AI-genererade feedbacken effektivt, vilket enligt dem berodde på hur feedbacken var formulerad. Eleverna upplevde ofta att feedbacken var generell, vilket gjorde att det var svårt för dem att veta exakt vad de behövde förbättra. Detta resultat ligger i linje med forskning av Jakobson (2019), som betonar att specifik kritik är mer gynnsam för elevernas förståelse av vilka revideringar som förväntas ske. Jakobson påpekar att specifik kritik som identifierar specifika områden för förbättring, är mer effektiv eftersom eleverna får vägledning om hur de kan utveckla sin text.

Majoriteten av debattartiklarna blev bättre efter revidering. Denna bedömning gjordes efter en jämförelseanalys av de ursprungliga versionerna och de revideringar som gjorts. Genom detta identifierades två typer av revideringar: de som hade positiv/viss positiv påverkan på textkvalitén samt de som hade negativ/ingen påverkan. I den senare gruppen som även benämnts som “mindre goda lösningar”, var sju av tjugofyra texter. Detta tyder på att

feedbacken från AI i de flesta fall faktiskt var givande för textens utveckling. I en studie om bedömning av en sammanfattning av en studentskriven uppsats av De Winter et al. (2023) visade det sig att rätt promptar kunde identifiera både stilistiska och innehållsmässiga fel. Detta gjorde även den AI-genererade feedbacken i deltagarnas debattartiklar och även om vissa förbättringspunkter var mer eller mindre relevanta, kunde eleverna fortfarande använda minst en punkt för att förbättra sin text.

I studien genomfördes ingen explicit responsdialog mellan mig och eleverna. Vissa elever hade svårigheter att tolka och applicera feedbacken, vilket kan ha berott på att feedbacken innehöll vissa begrepp som eleverna helt enkelt inte förstod. Detta trots att prompten satte explicita krav på responsens språknivå "Feedbacken ska vara på ett språk som är begripligt för en tonåring.". Denna språkliga krock kan ha lett till att elever inte uppfattade vissa kommentarer som relevanta eller istället valde "enklare" punkter att revidera som de inte behövde hjälp med att tolka eller förstå. Uppenbarligen saknades en gemensam förståelse kring de ändringar som behövdes och varför de var viktiga. Enligt Skolforskningsinstitutet (2015) och Jakobson (2019) är en responsdialog nödvändig för att elever ska kunna förstå och dra nytta av feedbacken.

Rudolph et al. (2023) uppskattar att det kommer att ta cirka tio år för skolväsendet att fullt ut anpassa sig till denna nya verklighet där AI är etablerat inom skolan. De betonar att det är nödvändigt för lärare att anpassa sina undervisnings- och bedömningsmetoder för att följa denna utveckling. Detta kan göras genom att bland annat implementera fortbildningsprogram för att hjälpa lärare förstå och effektivt använda AI-verktyg i skolkontext.

6.2 Didaktiska implikationer

Användningen av AI i formativ bedömning inom skrivundervisning innebär både fördelar och utmaningar. Nedan diskuteras de viktigaste aspekterna av dessa implikationer med grund i studiens resultat.

6.2.1 Effektiv resursanvändning

En av de främsta fördelarna med att använda AI för formativ bedömning är att snabbt producera omfattande feedback till eleverna. Detta frigör tid och lärarresurser som möjliggör att lärare kan fokusera på andra pedagogiska uppgifter, som planering av lektioner. Lärare kan på så sätt öka sin effektivitet och ge mer uppmärksamhet åt varje elevs individuella behov.

6.2.2 Utmaningar med effektivitet

En av utmaningarna som kommer med att använda AI-genererad feedback är att säkerställa att den är relevant och begriplig för eleverna. Feedbacken måste vara tydlig och konkret för att eleverna ska förstå vad som förväntas av dem och hur de kan utveckla sina texter. Det är viktigt att AI kan generera feedback som är anpassad till elevernas språkliga och kognitiva nivå för att den ska vara effektiv.

6.2.3 Självständigt lärande

När elever får omfattande formativ feedback måste eleven samtidigt ta ett större ansvar för sin egen inlärning och skrivutveckling. Om eleverna lär sig genom feedbacken att identifiera och åtgärda brister i sina texter, finns det potential att främja deras självständiga lärande.

6.2.4 Normering och kreativitet

Ytterligare en utmaning med att använda AI för bedömning är risken att den normerar skrivande och begränsar elevernas kreativitet. ChatGPT utgår tack vare sin träning ofta utifrån en strikt ram som fokuserar på specifika aspekter av textgenrer och skrivande, vilket betraktas vara "korrekt". Detta kan hindra elever från att utforska olika stilar och röster i sitt skrivande.

6.3 Metodreflektion

De metoder som har använts i studien har visat sig relevanta för att besvara forskningsfrågorna och därmed uppnå studiens syfte att bidra med kunskap om effekter av formativ bedömning producerad av artificiell intelligens i gymnasiets svenskämne. För att utveckla studiens metodik gör jag följande reflektioner. Prompten kan revideras ytterligare, så att den ger mer konkreta förslag på ändringar. I studien ger eleverna återkommande uttryck för att det ibland var diffust vad den AI-genererade kommentaren avsåg och på vilka sätt eleven därmed skulle utveckla texten. Dessutom noteras att den AI-genererade feedbacken varken registrerade eller kommenterade kraven för uppgiften. Till exempel stod det att elevtexterna skulle vara mellan 500–800 ord i prompten. Trots detta så fanns det ingen kommentar om textlängd i ChatGPTs respons. Detta hade behövts specificeras i prompten då det var information som ChatGPT inte fick något direktiv om vad den skulle göra med.

Vidare hade det varit gynnsamt att inkludera fler elever i studien, särskilt från

yrkesförberedande program. I studiens elevgrupp var andelen som gick på ett yrkesprogram för få för att göra några systematiska jämförelser mellan elevgrupper på yrkesprogram och högskoleförberedande program, totalt sex elever, varav två elever inte lämnade in sina texter i tid för att få AI-genererad feedback. En större och mer representativ grupp av elever hade kunnat ge en bredare insikt i elevernas uppfattningar när det gäller att implementera formativ feedback från ChatGPT i gymnasiets skrivundervisning.

En annan aspekt som hade kunnat förbättra studien är att använda en nyare version av ChatGPT. Istället för den använda gratisversionen (GPT-3.5) hade betalversionen GPT-4 kunnat användas som verktyg. En nyare version kan potentiellt ge mer precisa och relevanta svar, med tanke på att språkmodellen är en uppdaterad version och innehåller flera förbättringar och nya funktioner som skiljer den från den äldre versionen. Enligt OpenAI (n.d.b) har den senare modellen förbättrad prestanda, särskilt inom områden som tillförlitlighet, kreativitet och förmåga att hantera mer nyanserade instruktioner.

6.4 Fortsatt forskning

Resultaten i denna studie har ny kunskap, om än begränsad, om hur AI-genererad formativ feedback uppfattas och kan användas av elever. En större studie med fler deltagare hade kunnat ge mer omfattande resultat samt tydligare mönster. Det hade också varit intressant att inkludera deltagare från olika åldersgrupper och inte enbart gymnasiet för att se om det finns skillnader i effekter när det gäller att ta emot denna typ av feedback. Trots att denna studie hade formativa inslag, låg dess fokus på de specifika elevtexterna snarare än hur elevers skrivande kan utvecklas generellt. Studien kan därför inte erbjuda insikter om elevers generella skrivutveckling. För att undersöka dessa samband mellan AI-genererad feedback och skrivutveckling krävs longitudinella studier.

Vidare skulle forskning om att integrera AI som ett verktyg för bedömning kontinuerligt under elevens skrivprocess, inte bara som ett sista steg vara önskvärt. Då detta är ett outforskat område, behövs mer forskning för att förstå hur AI kan användas på ett funktionellt sätt för att stödja elevers skrivutveckling. Det hade även varit intressant om det gjordes en studie på de långsiktiga effekterna av AI-genererad feedback jämfört med traditionell lärarfeedback. Genom att undersöka hur användningen av AI-genererad feedback över tid påverkar elevernas skrivförmåga och inlärning hade en förståelse om det faktiskt är ett attraktivt och effektivt verktyg att implementera i skrivundervisning.

7 Referenslista

Black, P. & Wiliam, D. (2010). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan, 92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment (UCLA-CSIEP), 1*(2), 1–12.

Bloom, B. S. (1971). Mastery learning. I J. H. Block (Red.), *Mastery learning: Theory and practice*. (s. 47–63). Holt, Rinehart & Winston.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

De Winter, J. & Dodou, D. & Stienen, A. (2023). ChatGPT in Education: Empowering Educators through Methods for Recognition and Assessment. *Informatics. (10)*4, 87. https://doi.org/10.3390/informatics10040087

Duong, M., Thuy, M., Can Van, D. & Nguyen Van, H. (2024). The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. Sec. *Frontiers in Education (9),* 1–17. https://doi.org/10.3389/feduc.2024.1328769

Dysthe, O., Hertzberg, F. & Hoel, T.L. (2011). *Skriva för att lära: skrivande i högre utbildning.* (2., [rev.] uppl.) Lund: Studentlitteratur.

Holmes, W., Bialik, M. & Fadel, C. (2019). *Artificial intelligence in education: promises and implications for teaching and learning*. Boston, MA: The Center for Curriculum Redesign.

Faigley, L., & Witte, S. (1981). Analyzing Revision. *College Composition and Communication, 32*(4), 400–414. http://dx.doi.org/10.2307/356602

Fidan, M & Gencel, N. (2022). Supporting the Instructional Videos With Chatbot and Peer Feedback Mechanisms in Online Learning: The Effects on Learning Performance and Intrinsic Motivation. *Journal of Educational Computing Research. 0*(0) 1–26. https://doi.org/10.1177/07356331221077901.

Hattie, J. & Timperley, H. (2007). The power of feedback. I: *Review of Educational Research 77 (1)*, 81–112.

Hedman, E. (24 januari 2023). *Så kan AI förändra lärares arbete i grunden*. Vi Lärare. https://www.vilarare.se/nyheter/digitalisering/sa-kan-ai-forandra-larares-arbete-i-grunden/

Jakobson, L. (2019). *Skriftlig lärarrespons för vuxna nybörjare i svenska som andraspråk: teoretiska perspektiv, responspraktik och uppfattningar.* Diss. (sammanfattning) Göteborg: Göteborgs universitet, 2019. Göteborg.

Kronholm-Cederberg, A. (2009). *Skolans responskultur som skriftpraktik: gymnasisters berättelser om lärarens skriftliga respons på uppsatsen*. Diss. Vasa: Åbo Akademi, 2009. Vasa.

Lagerholm, P. (2024). *Språkvetenskapliga uppsatser.* (Tredje upplagan). Lund: Studentlitteratur.

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M. och López-Meneses, E. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers 12*(8), 153. https://doi.org/10.3390/computers12080153

OECD. (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. OECD Publishing. Paris. https://doi.org/10.1787/9789264007413-en

OpenAI. (n.d.a). *How can ChatGPT be used for assessment and feedback?* https://help.openai.com/en/articles/8313397-how-can-chatgpt-be-used-for-assess

ment-and-feedback

OpenAI. (n.d.b). *GPT-4 Research*. https://openai.com/index/gpt-4-research/

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342–362. https://doi.org/10.37074/jalt.2023.6.1.9

Sadler, D. Royce. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science 18*(2), 119–144.

Skolforskningsinstitutet (2018). *Feedback i skrivundervisningen*. [Solna]: Skolforskningsinstitutet.

Sok, S. och Heng, K. (2023). ChatGPT for Education and Research: A Review of Benefits and Risks. Cambodian Journal of Educational Research. (3)1, 110–121. https://doi.org/10.62037/cjer.2023.03.01.06

Vetenskapsrådet (2017). *God forskningssed* [Elektronisk resurs]. (Reviderad utgåva). Stockholm: Vetenskapsrådet.

World Economic Forum. (2023). *AI: Accelerating Students' Holistic Development in Teaching and Fulfilling*. Hämtad från https://www.weforum.org/agenda/2023/05/ai-accelerate-students-holistic-development-teaching-fulfilling/

Zirar, A. (2023). Exploring the impact of language models, such as ChatGPT, on student learning and assessment. *Review of Education (Oxford), 11*(3), 1–18. https://doi.org/10.1002/rev3.3433

8 Bilagor

Bilaga 1: Prompt inklusive deltagarnas uppgiftsbeskrivning till debattartikel

”Nedan finns det en uppgiftsbeskrivning för debattartikel i Svenska 1. Jag vill att du ger personlig feedback på: helhetsbedömning (Kommunikativ kvalitet – anpassning till syfte, mottagare och situation, självbärande text), innehåll & textanvändning (aktiviteter, innehållslig komplexitet), disposition & sammanhang (textbindning, inledning, avslutning, styckeindelning, styckemarkering, omfång), språk & stil (skriftspråksnormer, variation, välformulerat språk, situationsanpassning). Elevens debattartikel som du ska ge feedback på, klistrar jag in här nedan. Feedbacken ska vara formativ så både kvaliteter som är bra ska lyftas samt kvaliteter som kan utvecklas samt hur de ska/kan utvecklas. Ge konkreta/specifika exempel. Feedbacken ska vara på ett språk som är begripligt för en tonåring.

Uppgiftsbeskrivning: Du har själv valt ett ämne som intresserar dig och det är nu dags att börja skriva en debattartikel. Du ska skriva en övertygande text genom att formulera en tes och flera argument som stödjer denna tes. Ni har tillgång till en skrivmall här nedan samt en lathund som ligger uppe som material på Classroom. Din text ska också vara sammanhängande, begriplig och korrekt skriven och innehålla mellan 500-800 ord.

KRAV PÅ INNEHÅLL:
- En tydlig tes och tre argument för din tes samt ett motargument du bemöter
- En tydlig inledning och avslutning
- Ett korrekt skrivet språk som följer standardreglerna

Lycka till!”

Elevtext: *här har elevtext infogats*”

Bilaga 2: Bedömningsmatris för debattartikel Svenska 1

Bedömningsmatris – inlägg

Uppgift 2: Inför allas blickar

Svenska 1

Aspekter	F	E	C	A
Innehåll och kritisk läsning Aktiviteter Innehållslig komplexitet Bearbetning av textunderlag Källhantering		Eleven resonerar kring varför muntliga presentationer kan upplevas som obehagliga och diskuterar hur talängslan kan arbetas bort. Eleven presenterar något från Letmarks artikel. I sin text skiljer eleven på egna och andras tankar. Källhänvisningar finns om än inte helt fullständiga.	Innehållet är som helhet fylligt och relevant. Det som eleven presenterar från Letmarks artikel är rättvisande och väl valt. Källhänvisningen är infogad i texten.	Innehållet är nyanserat *eller* ger nya, relevanta perspektiv på ämnet eller i förhållande till Letmarks artikel.
Disposition och sammanhang Textbindning Inledning, avslutning Styckeindelning, styckemarkering Omfång		Texten är sammanhängande och begriplig, t.ex. genom referensbindning och enkla sammanhangssignaler. Omfånget är till viss del anpassat till situationen, dvs. texten är inte alltför kort.	Texten är tydligt strukturerad genom minst två av följande kännetecken: – varierade sammanhangssignaler – inledning och avslutning – rimlig styckeindelning som är konsekvent markerad.	Texten är väldisponerad genom följande kännetecken: – god balans mellan olika delar – effektiv inledning och avslutning – rimlig styckeindelning som är korrekt markerad.
Språk och stil Skriftspråksnormer Variation Välformulerat språk Situationsanpassning		Eleven följer i huvudsak skriftspråkets normer för språkriktighet, t.ex. i fråga om meningsbyggnad, stavning och skiljetecken.	Språket är varierat och delvis välformulerat. Stilen är anpassad till syftet, t.ex. genom variation mellan sakliga och värderande uttryck eller retoriska stilmedel såsom läsartilltal.	Språket är välformulerat, även om någon mindre lyckad formulering kan förekomma. Stilen är väl anpassad till syftet, t.ex. genom variation av meningstyper, uttrycksfulla formuleringar eller illustrativa förklaringar.
Helhetsbedömning Kommunikativ kvalitet – anpassning till syfte, mottagare och situation Självbärande text		Textens innehåll och form är till viss del anpassade till syfte, mottagare och kommunikationssituation. Texten fungerar efter viss bearbetning som ett självbärande inlägg på webbplatsen.	Textens innehåll och form är anpassade till syfte, mottagare och kommunikationssituation. Texten är självbärande, t.ex. genom ett mottagaranpassat innehåll och språk. Den fungerar som ett inlägg på webbplatsen.	Textens innehåll och form är väl anpassade till syfte, mottagare och kommunikationssituation. Texten fungerar väl som ett inlägg på webbplatsen.
Delprovsbetyg	**F**	**E**	**C**	**A**

EXEMPEL – Nationellt prov i svenska och svenska som andraspråk, kurs 1
Delprov C: skriftlig framställning – bedömning

Bilaga 3: Exempel på AI-genererad feedback på en stark debattartikel

Helhetsbedömning:

Din debattartikel är väl anpassad till syftet att övertyga läsaren om vikten av att tillhandahålla gratis hundbajspåsar till alla hundägare i Alingsås kommun. Du presenterar din tes tydligt och stöder den med övertygande argument och relevanta exempel. Din text är engagerande och lätt att följa, vilket gör den till en självbärande och övertygande text.

Innehåll & textanvändning:

Dina argument är väl utvecklade och stöds av tydliga exempel och fakta. Du visar en god förståelse för ämnet och framhäver viktiga konsekvenser av att inte plocka upp hundbajs. För att förbättra innehållet kan du inkludera ytterligare information om hur gratis hundbajspåsar har påverkat andra kommuner eller städer, vilket skulle stärka din argumentation ytterligare.

Disposition & sammanhang:

Texten har en tydlig struktur med en inledning, kropp och avslutning. Du använder styckeindelning effektivt för att organisera dina idéer, och din avslutning summerar övertygande dina argument. För att förbättra sammanhanget kan du använda övergångsfraser eller klargörande meningar för att smidigare koppla samman dina idéer och argument mellan olika delar av texten.

Språk & stil:

Ditt språk följer skriftspråksnormerna och är överlag välformulerat. Du visar en viss variation i ditt språk, men vissa delar av texten kan vara mer engagerande och levande för läsaren. Genom att använda mer varierade språkliga tekniker och stilistiska grepp, såsom retoriska frågor eller metaforer, kan du göra texten mer dynamisk och övertygande.

Exempel på förbättringar:

- Inkludera ytterligare information om tidigare framgångar: Genom att inkludera ytterligare information om hur gratis hundbajspåsar har påverkat andra kommuner eller städer, kan du stärka din argumentation och övertyga läsaren om effektiviteten av detta åtgärd. Detta skulle ge din text mer trovärdighet och övertygande kraft.

- Använd övergångsfraser för smidigare sammanhang: Genom att använda övergångsfraser eller klargörande meningar kan du smidigare koppla samman dina idéer och argument mellan olika delar av texten. Detta skulle förbättra sammanhanget och läsbarheten för din text.

- Använd mer varierade språkliga tekniker: Genom att använda mer varierade språkliga tekniker och stilistiska grepp kan du göra texten mer dynamisk och engagerande för läsaren. Du kan experimentera med olika typer av retoriska frågor, metaforer eller liknelser för att öka textens effektivitet och övertygande kraft.

Bilaga 4: Exempel på AI-genererad feedback på en svag debattartikel

Helhetsbedömning:
Din debattartikel är inriktad på att övertyga läsaren om vikten av godare skolmat för att förbättra elevernas arbetsförmåga och hälsa. Dock behöver texten förbättras för att bli mer självbärande genom att tydligare presentera och utveckla dina argument.

Innehåll & textanvändning:
Dina argument för att förbättra skolmaten är relevanta och ger stöd åt din tes. Du tar upp vikten av energi för elevernas arbetsförmåga, näringsinnehållet i skolmaten och minskningen av matsvinn. För att förbättra innehållet kan du utveckla dina argument ytterligare genom att inkludera exempel eller forskningsresultat som visar hur förändringar i skolmaten kan påverka elevernas hälsa och skolprestationer.

Disposition & sammanhang:
Du har en tydlig inledning där du presenterar ämnet och din tes. Du använder dig av flera argument för att stödja din ståndpunkt och avslutar artikeln genom att sammanfatta dina argument och betona vikten av förändring. För att förbättra dispositionen kan du arbeta på att markera övergångar mellan idéer tydligare för att underlätta för läsaren att följa din argumentation och skapa ett bättre sammanhang.

Språk & stil:
Ditt språk är överlag lättförståeligt och följer skriftspråksnormerna. För att förbättra språket och stilen kan du arbeta på att undvika grammatiska fel och variera din formulering för att göra texten mer engagerande och intressant att läsa.

Exempel på förbättringar:
- **Utveckla dina argument med exempel och forskning:** Genom att inkludera konkreta exempel eller forskningsresultat kan du göra dina argument mer övertygande och stärka din argumentation. Använd exempel som visar hur näringsriktig skolmat kan påverka elevernas hälsa och skolprestationer positivt.
- **Markera övergångar mellan idéer tydligare:** Använd övergångsfraser eller klargörande meningar för att binda ihop dina idéer och underlätta för läsaren att följa din argumentation. Tydliga övergångar bidrar till ett bättre sammanhang och en mer lättläst text.
- **Variera din formulering för att göra texten mer engagerande:** Genom att undvika upprepningar och variera din formulering kan du göra din text mer intressant och engagerande för läsaren. Använd olika ord och uttryck för att hålla läsarens intresse levande genom hela texten.

Bilaga 5: Intervjuguide

1. Hur kändes det att få AI-genererad feedback på era debattartiklar?
2. Hur skiljer sig er upplevelse av att få feedback från AI jämfört med att få feedback från en lärare?
3. Hur användbar och informativ upplevde ni feedbacken från AI vara?
4. Vilken feedback från AI kunde ni använda på egen hand för att förbättra era debattartiklar?
5. Vilken feedback från AI behövde ni hjälp av lärare för att tolka?
6. Kravet var att åtgärda minst två punkter av feedbacken, åtgärdade ni fler punkter? Varför/Varför inte?
7. Finns det några specifika aspekter av feedbacken från AI som ni uppskattade särskilt mycket?
8. Fanns det några svårigheter eller utmaningar med att ta till sig feedbacken från AI för att göra era revideringar?
9. Hur tycker ni att feedbacken från AI utvecklade kvaliteten på era debattartiklar?
10. Fanns det något specifikt område där ni önskar att AI kunde gett mer detaljerad feedback?
11. Har AI-genererad feedback bidragit till din skrivutveckling på något sätt?
12. Finns det några ytterligare tankar eller reflektioner ni vill dela om er erfarenhet av att ha fått AI-genererad feedback?