

Intelligent evaluering – på vej mod en model for systematisk evaluering af brugen af generativ AI



På skoleområdet er der stigende interesse for at anvende generativ AI (GenAI) såsom ChatGPT, der kan simulere evnen til skriftlig og visuel kommunikation om fag og sag. I artiklen behandles følgende problemstilling: Hvordan kan man systematisk evaluere brugen af GenAI for herved at styrke elevernes læringsbestræbelser?

I artiklen præsenteres en evalueringsmodel, som kan anvendes til at strukturere lærernes opmærksomhed på både tilsigtede og utilsigtede konsekvenser af brug af GenAI. Modellen er forskningsinformeret og rummer seks dimensioner af typiske lærerstyrede og elevcentrerede aktiviteter, der alle har relativ stor indflydelse på elevernes udvikling af faglig forståelse og deres baggrund for at tage stilling og handle. Anvendelse af den helhedsorienterede model kan komplettere nuværende tilgange til at evaluere feltet med henblik på at træffe informerede valg vedrørende hensigtsmæssig brug af GenAI i grundskolen.

Set i et lærerperspektiv omfatter modellen tre hovedtyper, nemlig initiering af læringsaktiviteter, løbende feedback og formativ vurdering af elevernes læringsbestræbelser. Set i et elevperspektiv omfatter den elevernes 1) udvikling af forforståelse af et fænomen, inden de begynder at arbejde selvstændigt med dette, 2) individuelle og kollaborative selvvirksomhed og 3) fremvisning af forståelse ved afslutningen af et forløb.

Nøgleord: formativ evaluering, evalueringsmodel, GenAI-anvendelse, grundskole

Indledning

Det har vakt opmærksomhed i samfundet generelt og i særdeleshed på uddannelsesområdet, at der er sket gennembrud i forskningen om intelligente systemer, forstået som digitale systemer, der kan simulere evnen til skriftlig og visuel kommunikation. Omkring 65 års forsknings- og udviklingsindsats bar frugt, da ChatGPT 3.0 blev lanceret i november 2022, og siden er det gået stærkt med at finde anvendelse for denne form for generativ AI (GenAI). Endda meget stærkt.

På uddannelsesområdet skabte lanceringen nærmest panik på grund af risiko for plagiat. Nu kunne man jo ikke længere vide, hvem der lavede store dele af elevernes skriftlige skolearbejde. Det har man imidlertid ikke kunnet længe, og den stigende anvendelse af GenAI har blot sat skub i overvejelserne over, hvilke former for daglig brug der kan give et retvisende billede af faglig indsigt og kunnen. Eksterne og interne udvalg på uddannelsesområdet har overvejet forskellige løsninger på dette dilemma. Det tegner til en blød landing, hvor elever kan bruge GenAI med alderssvarende indhold til emne- og projektarbejder og andre former for frie hjemmeopgaver, idet de er åbne om, hvordan de gør det.



Man kunne tro, at stormen herefter var afblæst. Men faktisk er den først lige er begyndt, for "alt dette skaber både nye muligheder og nye udfordringer" (Flyverbom et al. 2024: 4). I skrivende stund er det et åbent spørgsmål, hvilken indflydelse brugen af GenAI kan få på elevernes skoleengagement og udbytte af deres tiårige grundskoleforløb. Ved en sådan form for innovation er der så godt som altid nogle tilsigtede konsekvenser og nogle konsekvenser, som er uønskede (Rogers 1999). Problemet er, at vi ikke ved, hvori de består. Det gør det både svært at handle proaktivt, at maksimere fordelene ved læreres og elevers brug af GenAI – og minimere de uhensigtsmæssige følgevirkninger.

Man kan vælge at vente og se, hvad der sker, frem for at igangsætte en systematisk evaluering af dette. Denne strategi har tidligere været anvendt på uddannelsesområdet med meget uheldige følger. Jeg tænker her på brugen af sociale medier. Efter et årti er der skabt vished om, at omfangsrig brug af sociale medier har nogle alvorlige bivirkninger, og fokus er skiftet fra teknologioptimisme til overvejelser over forskellige former for begrænsninger og deciderede forbud.





Omkring ti årgange af børn og unge har i mellemtiden deltaget i et generationseksperiment, som kunne være undgået, hvis man på et tidligt tidspunkt, havde igangsat og gennemført evalueringer af, hvilke uhensigtsmæssige følgevirkninger der er ved omfangsrig brug af sociale medier.

Efter omkring ti års tøven trykker politikerne nu på panikknappen og går ind for begrænset skærm-tid. Denne form for reaktion kan sammenlignes med kostråd. Selv om vi godt ved, at noget er sundt, og noget er usundt, er det uholdbart at forbyde spisetid. Tilsvarende er det selvsagt ikke særlig nuanceret helt at forbyde skærmtid. I bestræbelserne på at undgå de skadelige følgevirkninger går man glip af en lang række fordele ved, at eleverne kan producere tekster og andre udtryk, få adgang til information, bruge læremidler og meget andet. Det kalder på, at man skelner klart mellem på den ene side struktureret skærmtid, som motiverer eleverne og fremmer deres læringsaktiviteter, og på den anden side skærmtid, som distraherer dem, og som i værste fald kan føre til mistrivsel og skabe afhængighed.

I artiklen behandles følgende problemstilling:

Hvordan kan man systematisk evaluere brugen af GenAI for herved at styrke elevernes læringsbestræbelser? Artiklen er disponeret således: Først formuleres en begrundelse for at evaluere brugen af GenAI, hvor hovedargumentet er, at sporene fra brugen af sociale medier skræmmer. Derpå præsenteres en forskningsinformeret evalueringsmodel, som kan henlede evaluatorernes opmærksomhed på nogle af de mest afgørende aspekter af implementeringen af GenAI i læringsmiljøer. Artiklen afsluttes med en diskussion og en konklusion.

Hvis vi vil undgå at gentage denne uheldige historie i sammenhæng med GenAI, er det på tide at igangsætte systematisk evaluering.

Hensigten i denne artikel er at give et forskningsinformeret bud på, hvad aktører på uddannelsesområdet med fordel kan fokusere på. I det følgende vil jeg præsentere en evalueringsmodel, som kan bruges, når man på den enkelte skole eller årgang vil evaluere brugen af GenAI med henblik på at styrke fordelene og mindske ulemperne.

Sporene skræmmer

Camilla Mehlsen (2024) har skrevet en bog i den hensigt at "åbne den sorte boks til vores nærmeste digitale mediehistorie," og indholdet er ikke rar læsning. Det fremgår, at børn og unge er blevet udsat for et generationseksperiment uden sidestykke, og aldrig før i historien har så mange rapporteret om udfordringer med at fastholde opmærksomheden i samvær, hvor man mødes ansigt til ansigt med familie og venner samt ved fysiske aktiviteter og skolearbejde. Tilsvarende har der ikke tidligere været så stor en andel af unge, som rapporterer selvskade. Mehlsen omtaler også en teori (formuleret af Leon Festinger) om, at en grundlæggende drift hos alle mennesker går ud på at sammenligne sig med andre. Hvis man bestandigt bedømmer sig selv og sin formåen i forhold til andre, kan det fremme udviklingen af en præstationskultur, "en slags omsiggribende bedømmelses- og konkurrencekultur" (Sørensen et al. 2024).

"Nævnes skal også, at teenagepigens brug af sociale medier foranlediger et øget omfang af mentale lidelser såsom angst og depression, ligesom der er sammenhæng mellem henholdsvis denne brug og stigningen i omfanget af andre skadevirkninger såsom afhængighed, søvnunderskud og nedsat evne til at være opmærksom."

(Haidt 2024)

Nicklas Brendborg (2024) har skrevet en bog om superstimuli, og brugen af sociale medier anses for at være en af de mest udbredte af disse. En følgevirkning betegnes "opmærksomhedstyveri", der indebærer, at sociale medier får opmærksomhed i mange af døgnets vågne timer.





Det skyldes, at man som bruger kan blive afhængig af indhold, der er designet til at virke fastholdende og superstimulerende. Ikke overraskende tyder resultat af empirisk forskning inden for dette felt derfor på øget mentalt velvære, hvis man mindsker denne afhængighed.

For eksempel er der tegn på øget velvære, hvis familier reducerer deres skærmbrug i fritiden til mindre end 3 timer om ugen for hver person (Pederesen et al. 2022). For grundskoleelever, som er i fokus i denne artikel, tyder resultater af forskningen på, at ustruktureret brug af sociale medier i højst 30 minutter dagligt kan medvirke til en betydelig forbedring af elevernes velbefindende (Hunt et al. 2018).

Det kan sammenholdes med, hvor megen fritid eleverne i dag bruger på sociale medier og digitale spil. Tidsforbruget har været stigende siden lanceringen af smartphones (Brendborg 2024). En undersøgelse fem-seks år herefter viste, at relativt mange 11-15-årige piger hele tiden overvejede, hvornår de næste gang kunne gå på sociale medier, og for en tredjedel af drengene i samme aldersgruppe var tidsforbruget på digitale spil over fire timer dagligt (Rasmussen et al. 2019). I nutiden bruger danske elever faktisk mere tid foran digitale skærme såsom mobiltelefoner, tablets, computere og fjernsyn end elever i andre lande (Ottosen & Andreasen 2020). For afgangselever i folkeskolen er det gennemsnitlige tidsforbrug steget til fire-fem timer om dagen (Madsen et al. 2023).

Der ser ud til at være en sammenhæng mellem stigende anvendelse af sociale medier i elevernes fritid og rapportering af psykiske udfordringer. Det er i hvert fald tilfældet, hvis man sammenholder kurver over elevernes forbrug af sociale medier og deres selvrapporterede psykiske udfordringer, idet udviklingen er helt parallel (Brendborg 2024). Risikoen for selvoplevet stress blandt børn og unge kan øges i situationer med "afhængighed af spil," stort "tidsforbrug på sociale medier/internettet" og "mindre søvn pga. skærmtid" (Kierkegaard et al. 2020).



"For nogle elever kan sådanne medievaner også "skabe bekymringer, give en oplevelse af ensomhed, lavt selvværd eller en oplevelse af at være uden for et fællesskab."

(Sundhedsstyrelsen 2023: 9)

Oplevelsen af afhængighed af sociale medier og et højt tidsforbrug i anvendelsen af dem øger i det hele taget risikoen for, at eleverne oplever koncentrationsbesvær (Sundhedsstyrelsen 2023). For eksempel oplever de, at det er svært at koncentrere sig om fagligt indhold, fordi de så at sige bliver fanget af fastholdelsesmekanismer og afhængighedsskabende indhold på sociale medier.

Det siger næsten sig selv, at det udfordrer elevernes faglige læsning, hvis læsetiden er knap, og opmærksomheden er fragmenteret. På dette område er der endnu en kurve, som skal knækkes. I 2021 er andelen af elever på 4. klassetrin med stor tiltro til egen læseforståelse nemlig signifikant mindre end i 2016 – og i denne periode er afstanden mellem de svageste og dygtigste læsere også blevet større (Fougst 2023). Den stigende andel af læseusikre elever risikerer at miste deltagelsesmuligheder i skole og fritid på kortere sigt samt uddannelses- og livsmuligheder på længere sigt.

Selv om sammenhængen mellem forbrug af sociale medier og psykiske udfordringer er blevet påvist gennem en årrække, er det først i de senere år, at man på skolerne for alvor har fulgt op på denne sammenhæng. For at bremse en udvikling, hvor generationseksperimentet kan få mere og mere alvorlige følger for en del elever, er man på flere og flere skoler begyndt at indføre begrænsninger i forbruget af sociale medier og digitale spil – og man bestræber sig på at efterkomme Sundhedsstyrelsens vejledninger på området (2023). Dette gennemgribende holdningsskift kan – som foreslået af Mehlsen (2024) – sammenlignes med "tobaksøjeblikket", hvor der blev konsensus om restriktioner for at forebygge tobakkens sundhedsskadelige virkninger.



Dette skift fra en holdning præget af teknologioptimisme og laissez faire og til en holdning præget af betænkelighed i forhold til børn og unges brug af techgiganternes produkter, såsom digitale medier, platforme og spil, kan begrundes i resultater af evalueringer og forskning om u hensigtsmæssige følgevirkninger. Det har blandt andet sat øget fokus på perioder med mediestilhed.

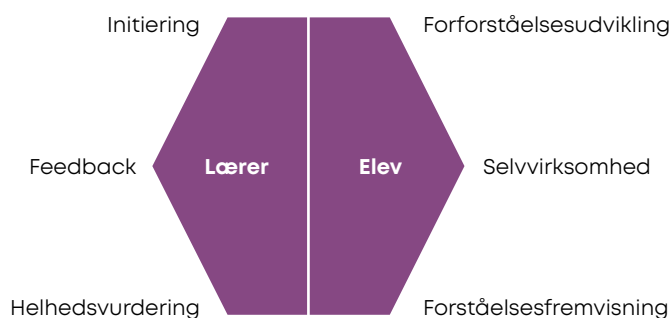
Som før omtalt er det for unuanceret at forbyde al skærmtid. I stedet kan man søge at skabe en passende balance mellem dels struktureret brug af digital teknologi til lokalisering, behandling og formidling af tekst, tal, billeder og lyd, dels alle de før omtalte u hensigtsmæssige følgevirkninger.

Evaluering af elevernes brug af GenAI

Med henblik på at undgå en gentagelse af generationseksperimentet med sociale medier – og i stedet fremme tilsigtede og mindske utilsigtede konsekvenser af brug af GenAI – vil jeg som nævnt anbefale, at man løbende evaluerer denne brug. Ved at gøre det nu kan man skabe et bedre grundlag for at udnytte potentialerne for læring og mindske eventuelle udfordringer på dette felt. Denne anbefaling er på linje med anbefalinger i den internationale litteratur på området. For eksempel anbefaler det britiske undervisningsministerium, at man skaber vished gennem løbende evalueringer af brugen af GenAI i skolerne, og ministeriet overvejer at støtte skolerne i dette både økonomisk og på anden måde (DfE 2024).

For mig at se er en løbende evaluering temmelig oplagt. Hvis man vil vide, hvilke erfaringer skolens primære aktører gør sig med brug af GenAI, hvorfor så ikke involvere dem? De primære evaluører omfatter i denne sammenhæng lærere og ledere på folkeskoleområdet, idet de kan påtage sig at tilrettelægge og gennemføre løbende evalueringer af brugen af GenAI. Det omfatter også involvering af eleverne i de løbende evalueringer, selv om det kan bevirke, at det tager lidt længere tid, end hvis man kun evaluerer på grundlag af lærernes og ledernes erfaringer og refleksioner.

Til brug for løbende evalueringer har jeg som før omtalt udviklet en forskningsinformeret model, som er illustreret i figur 1. Det fremgår af figuren, at modellen omfatter et lærer- og et elevperspektiv, og at der for begge er tre centrale dimensioner. Hensigten i det følgende er at beskrive hver af disse dimensioner nærmere.



Figur 1. Model til evaluering af lærerstyrede og elevcentrede aktiviteter med GenAI.

Lærerstyrede aktiviteter

Evalueringsmodellen omfatter aktiviteter, som en lærer typisk igangsætter og gennemfører inden for forskellige fagområder i samarbejde med en gruppe elever. Som det fremgår af figur 1, omfatter den tre hovedtyper, nemlig initiering af læringsaktiviteter, konstruktiv feedback og helhedsorienteret vurdering af elevernes læringsbestræbelser.

Dimensionen *initiering* skal forstås bredt som indføring af eleverne i faglige emner, begreber og arbejdsmetoder samt besvarelse af spørgsmål om undervisningens *hvad*, altså hvad eleverne skal lære, *hvorfor*, altså formål og mål med dette, og *hvordan*, herunder organisering og arbejdsmetoder. En præsentation af indholdsviden, der er ny for eleverne, kan siges at være vellykket, når alle elever "forstår og husker dette indhold" og "kan se pointen med at sætte sig ind i og anvende det" (Illeris 2017: 22).

Initiering omfatter derudover instruktion af eleverne til selvstændige læringsaktiviteter. Mere lavpraktisk er hensigten selvsagt at begrænse mængden af deres tvivlsspørgsmål såsom "hvad skal vi gøre nu?" og "hvorfor og hvordan gør vi det?" Elevernes læringsudbytte af klare udmeldinger om kommende aktiviteter kan bl.a. vurderes på grundlag af *effektstørrelser*. Ifølge Hatties sammenfatning af en lang række empiriske undersøgelser er den gennemsnitlige effektstørrelse af forskellige læringsfremmende tiltag 0,4 (Hattie 2009), men for initiering er den 0,75 (Hattie 2018), altså næsten dobbelt så høj.

Når eleverne skal arbejde på egen hånd, kan de stå over for to udfordringer. Først skal de forstå indholdsviden, der er ny for dem, og derefter skal de anvende den nyerhvervede forståelse til opgave- eller problemløsning eller andre former for selvstændigt arbejde. For nogle elever kan en sådan dobbeltudfordring begrænse deres lærelyst og udbyttet af deres selvvirksomhed med videnindholdet. Derfor er det i mange tilfælde ikke optimalt at præsentere eleverne for opgaver eller problemstillinger, som de skal håndtere på egen hånd uden at give forudgående vejledning.

"Vejledning i fællesskab af en gruppe elever, hvilket også kan betegnes guidende undervisning, har en relativt stor indflydelse på elevernes læringsresultater med en gennemsnitlig effektstørrelse på 0,60."

(Hattie 2018)

Vejledning kan i særdeleshed hjælpe eleverne med at afgrænse kommende arbejdsopgaver og introducere dem til, hvordan de kan opfylde målene for deres selvstændige arbejder. For eksempel er den generelle indflydelse fra "problemløsningsundervisning" på elevernes læringsudbytte relativt høj med en effektstørrelse på 0,68 (Hattie 2018). Ved problembaseret læring *uden* vejledning er den gennemsnitlige effektstørrelse derimod blot 0,26 og dermed væsentligt under den gennemsnitlige effektstørrelse (Hattie 2018). Der er også udsigt til en reduktion af det faglige udbytte, hvis lærerne helt overlader kontrollen over læringen til eleverne selv. Faktisk er effektstørrelsen af "deltagerkontrol over læring" i gennemsnit blot 0,02 (Hattie 2018).

Der er således flere argumenter for, at læreren skal vejlede eleverne, så de alle forstår meningen med deres kommende læringsaktiviteter, herunder ikke mindst deres kommende selvvirksomhed. Der er indtil videre sparsom viden om, hvorvidt og i hvilket omfang lærerne kan drage nytte af at bruge GenAI til dette formål, men en undersøgelse blandt lærere i grundskolen i England, Skotland og Wales tyder på, at det hyppigst sker for at frembringe undervisningsmateriale til eleverne (DfE 2024). For eksempel kan lærerne vælge at "bruge GenAI til at skabe skræddersyede læringsmaterialer og aktiviteter, der imødekommer elevernes faglige interesser" (Labadze et al. 2023: 14).

For at lette arbejdet med denne type rutineopgave kan lærerne lave et udkast og derefter finpudse det under indtryk af kommentarer fra et GenAI-system. Ved at spørge, om indholdet er klart og præcist formuleret, velstruktureret, tilstrækkeligt detaljeret samt relevant og engagerende, kan lærerne få inspiration til at forbedre indholdet af deres udmeldinger til eleverne. Med henblik på at lave de førnævnte skræddersyede præsentationer og materialer kan lærerne angive en kontekst og eksempelvis skrive: "Jeg er lærer i fag x og vil skrive en meddelelse, som kan forstås af elever på trin y i grundskolen." Desuden kan de specificere elevernes faglige forudsætninger og endda specificere nogle af elevernes tidligere læringserfaringer inden for et givet fagområde.

Dertil kommer, at fælles drøftelser, hvor indholdet er tilpasset elevernes forudsætninger, kan reducere omfanget af kognitive belastninger og skabe optimale "zoner for nærmeste udvikling" (Ferguson et al. 2022). Sådanne drøftelser kan hjælpe eleverne til at forstå, hvad det er hensigten, at de lærer at mestre, og øge deres autonomi i efterfølgende, selvstændige arbejder. I særdeleshed kan fælles drøftelser af faglige emner, begreber og arbejdsmetoder medvirke til at involvere og engagere eleverne i planlægning af egen selvvirksomhed. Indflydelsen fra fælles drøftelser i klasserummet på elevernes læringsresultater er faktisk relativt høj med en effektstørrelse på 0,82, altså dobbelt så høj som den gennemsnitlige effektstørrelse (Hattie 2018).

Efter en fælles drøftelse kan læreren opfordre eleverne til at bruge GenAI, hvis de har behov for yderligere vejledning. Herved kan de få supplerende oplysninger og svar på tvivlsspørgsmål, som kan øge deres muligheder for at engagere sig i læreprocesser, hvor der differentieres med hensyn til bredde, dybde og tempo i arbejdet med fagligt stof, og dette kan gøre det lettere for læreren at give differentieret vejledning.

Dimensionen betegnet feedback i figur 1 skal forstås bredt som situationer, hvor lærere giver løbende tilbagemeldinger til elever, individuelt eller i grupper, i den hensigt at fremme elevernes engagement og faglige udbytte (Webb et al. 2018). Når lærere giver feedback under eller umiddelbart efter en læringsaktivitet, er indflydelsen på elevernes læringsudbytte i gennemsnit relativt høj med en effektstørrelse på 0,75 (Black & William 1998).

Da eleverne har forskellige forudsætninger, må lærerne give differentieret feedback, der er afpasset i forhold til disse forudsætninger; men dette kan være svært at gøre i tilstrækkeligt omfang, når mange elever arbejder selvstændigt. Lærerne kan derfor opfordre eleverne til også at få feedback ved brug af GenAI, når de har behov for det. Rettidig brug af GenAI til dette formål kan være en fordel for eleverne og gøre feedback mere overkommelig for lærerne.

Dimensionen betegnet *helhedsvurdering* i figur 1 vedrører situationer, hvor lærerne indhenter og

analyserer information om elevernes læreprocesser og udbytte heraf. Særligt kan lærerne vurdere, om eleverne er aktive og parate til at lære, samt om de er passende grupperet og samarbejder på måder, der understøtter deres læring (Bower et al. 2015). I den forbindelse må lærerne være opmærksomme på elever, som snyder sig selv for læringsoplevelser. Principielt kan eleverne undlade selv at lave væsentlige dele af deres hjemmeopgaver, projekter og lignende, fordi de let kan få et GenAI-system til at lave dem (Hørløv 2023). Lærerne må derfor søge at vurdere både elevernes faglige aktiviteter og udbytte heraf.

Lærerne kan supplere sådanne former for helhedsvurderinger ved at opfordre eleverne til at bruge GenAI til selv at evaluere deres forståelse af indholdsviden. Det betyder, at nogle elever kan interagere med et GenAI-system, mens deres lærer vurderer andre elevers indsats og produkter. Læreren kan i så fald løbende vurdere, om denne form for komplementær vurdering reelt styrker elevernes autonomi og engagement.

Elevecentrerede aktiviteter

De elevrelaterede dele af evalueringsmodellen i figur 1 vedrører elevernes læreprocesser, hvorigennem de udvikler en ny forståelse af et fænomen. Der er tre hoveddele, nemlig:

- forforståelsesudvikling – elevernes udvikling af forforståelse og afgrænsning af et fænomen, inden de begynder at arbejde selvstændigt med dette
- selvvirksomhed – elevernes individuelle og kollaborative bestræbelser på at forstå fænomenet
- forståelsesfremvisning – elevernes fremstilling af deres forståelse af fænomenet ved afslutningen af et forløb.

I modellen henviser det relativt brede begreb forforståelse nærmere bestemt til de former for forståelse, som eleverne i første omgang udvikler af indholdet i en kommende læringsaktivitet. En elevs forforståelse kan også betegnes som en "for-dom". Det vil sige en dom, som eleven fælder på forhånd, og som ifølge hermeneutikken skaber en samlet horisont, hvorfra alt det følgende bliver

tolket (Gadamer 2000). I dette tilfælde omfatter det fordomme om læringsindholdets relevans og sværhedsgrad samt forløbets organisering og arbejdsformer. Ved at arbejde bevidst med forforståelse kan læreren opnå, at eleverne udvikler realistiske forventninger til egne læringsprocesser, som gradvist resulterer i opnåelsen af den tilsigtede forståelse (Koper & Bennett 2008).

”I mange tilfælde præsenterer læreren selv intentionerne med et kommende læringsforløb eller henleder elevernes opmærksomhed på skriftligt materiale med oplysninger om sådanne formål. Indflydelsen fra denne form for forforståelse er relativt høj, idet den generelle effektstørrelse af at eksplicitere mål (til forskel fra ingen mål at have) er 0,68.”

(Hattie 2018)

Som tidligere omtalt skal sådanne intentioner tilpasses elevernes læringsmæssige forudsætninger, herunder deres nuværende viden og færdigheder. Hvis eleverne oplever, at de bliver fagligt udfordret i passende grad, er effektstørrelsen i gennemsnit 0,59 (Hattie 2018). Derfor kan læreren og eleverne med fordel drøfte, hvad eleverne på forhånd ved om ny indholdsviden, og relatere dette til den intenderede forståelse. Faktisk er den gennemsnitlige effektstørrelse af strategier for at integrere ny viden med tidligere viden forholdsvis høj, nemlig 0,93 (Hattie 2018).

En anden dimension i evalueringsmodellen, som ses i figur 1, vedrører som nævnt *selvvirksomhed*, hvor hensigten blandt andet er, at eleverne udvikler viden og færdigheder ved at engagere sig i deres læringsaktiviteter. Det har almindeligvis positiv indflydelse på deres læringsudbytte, at de deltager aktivt (Deslauriers et al. 2019). For eksempel tyder forskningsresultater inden for STEM-discipliner på, at eleverne lærer bedst, når de er aktive i en fagligt orienteret praksis (Freeman et al. 2014). De kan engagere sig i at læse, reflektere, diskutere, udforske og problemløse samt lave

projektarbejde, der er egnet til at udvikle deres forståelse af indholdsviden.

I forbindelse hermed kan eleverne bruge GenAI i henhold til deres individuelle behov. Hvis de indtaster oplysninger om egne opgaver, kan de anmode om output, der er tilpasset deres faglige forudsætninger og læringsmæssige intentioner. De kan derefter reflektere over dette output og bruge det til at vurdere og efter behov forbedre deres skriftlige opgaver eller mundtlige præsentationer. Vi kan for eksempel forestille os, at nogle elever laver et projekt på et bestemt niveau inden for et bestemt fagområde. Hvis de benytter et GenAI-system, kan de indledningsvist prompte, at de er elever ”i grundskolen på niveau x,” og at de skal løse ”en opgave i x om emne y” eller lave ”et emnearbejde om z.” De kan derefter indsætte opgaven og deres udkast til løsning og derefter spørge: ”Kan jeg få feedback?”



På denne måde kan feedback, der er tilpasset elevernes læringsintentioner, hjælpe dem med at reflektere over det lærte og i det hele taget medvirke til at udvikle deres faglige forståelse. Den lette adgang til sådanne former for feedback kan især styrke deres muligheder for selv at vurdere, om deres forståelse af indholdsviden er fyldestgørende eller mangelfuld i forhold til på forhånd givne, alderssvarende intentioner for deres udvikling af forståelse. Ved at spørge med deres egne ord og måske mere end én gang kan de få output, som de fuldt ud forstår, og som styrker deres forståelse af faglige emner, begreber og arbejdsmetoder. Der er forskningsmæssigt belæg for, at sådanne former for elevinitieret "evaluering og refleksion" har stor indflydelse på elevernes læringsresultater med en gennemsnitlig effektstørrelse på 0,75 (Hattie 2018).

Desuden er der forskningsmæssigt belæg for, at elevernes motivation for at lære afhænger af, hvordan de værdsætter deres nuværende læringsaktiviteter (Pintrich & Schunk 2002). Den gennemsnitlige effektstørrelse af "opfattet opgaveværdi" er 0,46 (Hattie 2018). Som nævnt afhænger elevernes motivation også af deres tro på, at de vil være i stand til at gennemføre kommende læringsaktiviteter. Når de opnår positive mestringsoplevelser, kan det øge deres oplevede self-efficacy. Effektstørrelsen af at arbejde bevidst med styrkelse af elevernes self-efficacy er i gennemsnit relativt høj sammenlignet med andre former for indsatser, nemlig 0,92 (Hattie 2018).

Under arbejdet i grupper kan eleverne blandt andet få gavn af at give og modtage peer feedback (Webb et al. 2018). Sådanne samarbejdsformer, som medvirker til, at faglige udtryk bliver en del af elevernes aktive ordforråd, kan organiseres på mange måder (Kagan & Stenlev 2007). I gennemsnit er effektstørrelsen af et skift i fokus fra individuel læring til kooperativ læring 0,55 (Hattie 2018).

Den dimensionen i evalueringsmodellen, der betegnes forståelse, omfatter den *forståelse*, som eleverne opnår under et læringsforløb, og som de kan fremvise ved afslutningen af forløbet. Eleverne udvikler typisk en sådan forståelse ved at anlægge fagrelaterede synsvinkler (Gardner 2000). Da dette kræver fortolkning af faglig viden, er denne

form for læringsbestræbelse mere krævende, end hvis eleverne blot søger at reproducere en sådan viden (Illeris 2017).

Elever, der vil fremvise deres forståelse af indholdsviden mundtligt eller i form af skriftlige opgavebesvarelser, kan anvende et GenAI-system som "co-editor" (Haug & Haug 2023). For eksempel kan de formulere et udkast og få GenAI-forslag til forbedringer af indhold og form, inden de reviderer og formidler deres arbejder. Hermed kan de søge at styrke deres kommunikation i mundtlige oplæg, skærmpresentationer og skriftlige afleveringer. Både højt og lavt præsterende elever kan drage nytte af dette, og der er forskningsmæssigt belæg for, at det kan reducere forskelle i elevernes faglige præstationer (Winn 2023).

Diskussion

I denne artikel foreslås en udbygning af skolernes hidtidige evalueringspraksis med henblik på at skabe hensigtsmæssige rammer for lærernes og elevernes brug af GenAI. Grundskolens parter må træffe valg med hensyn til GenAI-anvendelsens hvad, hvordan og hvorfor (Elstad 2023). Hensigten med artiklen er at medvirke til at skabe et informeret grundlag for lærernes og elevernes proaktive valg af fremtidige anvendelsessituationer.

Hensigten er ikke at overflødiggøre hidtidigt anvendte evalueringsmodeller eller -redskaber, men at komplettere den hidtidige brug af sådanne for at strukturere deres opmærksomhed på brug af GenAI. Man kan skelne mellem de obligatoriske evalueringsredskaber og de redskaber, som skolerne selv vælger at bruge:

Skolerne bruger typisk en del redskaber til at følge henholdsvis trivsel, alsidig og faglig udvikling. Det gælder både folkeskoler og specialskoler. De obligatoriske redskaber som nationale test, fraværsregistreringer og karakterer suppleres således med redskaber, som skolerne selv vælger at bruge, hvilket indikerer, at de obligatoriske redskaber ikke er tilstrækkelige til formålet (Pedersen et al. 2018: 5).

Den foreslåede model for evaluering af brugen af GenAI er ment som et middel til at styrke opmærksomheden på denne brug, fremme de tilsigtede konsekvenser og reducere de utilsigtede konsekvenser i tilknytning hertil.

Der findes mange modeller og redskaber til evaluering, hvoraf langt fra alle er videnskabeligt dokumenteret.

Som nævnt er hensigten i denne artikel at præsentere en model med seks hovedtyper af aktiviteter, hvor der er forskningsmæssigt belæg for, at de har relativt stor indflydelse på skolerelaterede forbedringstiltag.

I forbindelse med skoleudvikling anvender man i nogle tilfælde SMTTE-modellen (Håstein 2017). Den er kendetegnet ved, at man kigger efter konkrete tegn på forbedringer og målopfyldelse. Det bud på en evalueringsmodel, som jeg har præsenteret i det foregående, kan bruges som et redskab til at kvalificere dette arbejde, idet modellen omfatter seks hovedtyper af tegn, der alle har relativt stor indflydelse på elevernes faglige udvikling.

Modellen vedrører ikke processer i tilknytning til GenAI-implementering i læringsmiljøer, men helhedsorienteret evaluering af brugen af GenAI i lærer- og elevperspektiv. Med andre ord er planlægning eller gennemførelse af GenAI-aktiviteter ikke i fokus som sådan; det er analyse af værdien heraf for dels lærernes arbejde og professionelle udvikling, dels elevernes motivation og læreprocesser. Modellen er parallel frem for sekventiel, idet de seks dimensioner betegner aktiviteter, der kan forekomme sideløbende. For eksempel forekommer nogle af de lærerstyrede aktiviteter som oftest sideløbende med nogle af de elevstyrede aktiviteter. Det giver mulighed for, at lærerne kan bruge GenAI til løsning af rutineopgaver, og eleverne kan bruge GenAI i forbindelse med deres læreprocesser (Elstad 2023).

En anden model, som har opnået en vis udbredelse på skoleområdet, er modellen "Læringsmiljø og

pædagogisk analyse" (også kaldet LP-modellen). Det er en sekventiel model, hvor man først identificerer udfordringer i læringsmiljøet. Derefter indhenter og analyserer man information om, hvad der skaber disse udfordringer, inden man udvikler, gennemfører og evaluerer tiltag for at reducere eller helt fjerne dem (Andresen et al. 2009). Det bud på en evalueringsmodel, som jeg har præsenteret i det foregående, kan komplettere en sådan indsats ved at udpege seks opmærksomhedspunkter, hvor anvendelse af GenAI har relativt stor indflydelse på elevernes faglige udvikling og lærelyst. Det er punkter, hvor det er vigtigt at være opmærksom på potentielle udfordringer og tiltag for at sikre, at alle så vidt muligt får gavn af adgangen til GenAI. Midlet hertil er som nævnt at strukturere opmærksomheden på denne brug.

Den hidtidige debat har budt på mange spekulationer om konsekvenserne på godt og ondt. Skolerne kan dog ikke vente på at få indsigt i de reelle konsekvenser, men må her og nu søge at få vished om, at der ikke er vedvarende udfordringer, som lærere og elever har vanskeligt ved at håndtere. Ligesom skolerne selvfølgelig må kende og søge at realisere de potentielle fordele, for eksempel at lærerne får flere handlemuligheder, og eleverne bliver mere selvhjulpne og får større tro på egen formåen i forhold til deres skolearbejde.

De seks dimensioner i den præsenterede evalueringsmodel er identificeret ved analyse af resultater af empirisk forskning. Som målestok er anvendt resultater, der bliver udtrykt som effektstørrelser. De bliver beregnet som et gennemsnit af indflydelsen af forskellige tiltag på elevernes læring. Dette er sket på tværs af undersøgelser, hvor et relativt stort antal lærere og elever fra forskellige skoler har deltaget. Selve forskningsmetoden kan diskuteres, men det falder uden for rammerne af denne artikel. Det skal blot nævnes, at Vembye (2022) har lavet en kritisk-konstruktiv analyse af fortællinger om resultater og statistiske beregninger, hvad angår anvendelse af metoden.

Det skal også nævnes, at et gennemsnit af indflydelser, som er opnået i andre kontekster end "tirsdag i tredje time i en bestemt klasse på en bestemt skole" ikke er nogen garanti for, at den samme indflydelse vil forekomme der. Der er kun én vej, hvis lærerne vil have overblik over indsatsen og mulighed for at handle proaktivt, og det er selv at evaluere brugen af GenAI i læringsøjemed. Lærerne kan især rette opmærksomheden mod, hvilke forudsætninger de selv og eleverne skal udvikle for at anvende GenAI på hensigtsmæssige og ansvarlige måder – og om eleverne rent faktisk gør det. Hermed kan de modvirke tendenser til, at nogle bliver efterladt på perronen uden mulighed for at realisere potentialerne i en velovervejet anvendelse af GenAI – for at forbedre lærernes arbejdsvilkår og elevernes læringsvilkår.

Konklusion

Der er en stigende interesse for at anvende GenAI på skoleområdet. Det skaber et behov for, at man tilvejebringer lokal viden gennem rutinemæssige evalueringer af styrker og svagheder ved anvendelsen. I denne artikel er følgende problemstilling blevet behandlet: Hvordan kan man systematisk evaluere brugen af GenAI for herved at styrke elevernes læringsbestræbelser?

Artiklen giver et bud på en evalueringsmodel, som kan supplere den hidtidige evalueringsindsats i grundskolen. Modellen er forskningsinformeret og rummer seks dimensioner, som alle har relativt stor indflydelse på elevernes faglige udvikling. Hensigten med at anvende den proaktivt er blandt andet at undgå en gentagelse af generationseksperimentet, hvad angår sociale medier og digitale spil. Det gik i mange år under radaren i skolen – med deraf følgende distraktion og andre skadevirkninger.

Anledningen til udvikling af modellen er den oftestenlige adgang – fra ultimo 2022 – til generativ AI (GenAI), eksempelvis ChatGPT. Det har ændret vilkårene for mange former for selvstændige elevarbejder i grundskolen. Grænsen mellem brug og misbrug af GenAI er hårfin, fordi eleverne på den ene side kan få inspiration, tekstforslag, svar på tvivlsspørgsmål og feedback på egne tekstudkast, mens de på den anden side kan snyde sig selv

for læringsoplevelser, hvis de overlader det til et GenAI-system at lave deres skolearbejde.

Tilsvarende er kravene til en løbende evaluering på dette felt blevet skærpet, fordi det kan være særdeles svært for lærerne at afgøre, om det gavner eleverne at bruge GenAI i deres skolearbejde, eller det modsatte er tilfældet. I artiklen præsenteres derfor en forskningsinformeret evalueringsmodel, som er udviklet med særligt henblik på at strukturere lærernes opmærksomhed på både tilsigtede og utilsigtede konsekvenser ved GenAI-brug i grundskolen.

I et lærerperspektiv omfatter den tre hovedtyper, nemlig initiering af læringsaktiviteter, konstruktiv feedback og formativ vurdering af elevernes læringsbestræbelser. I et elevperspektiv omfatter den elevernes 1) udvikling af forforståelse af et fænomen, inden de begynder at arbejde selvstændigt med dette, 2) individuel og kollaborativ selvvirksomhed og 3) fremvisning af forståelse ved afslutningen af et forløb.

Litteratur

- Andresen, B.B., Egelund, N. & Nordahl, T. (2009). LP-modellen – forskningsindsats og første resultat. *Pædagogisk Psykologisk Tidsskrift*, 46(6), 386-394.
- Bower, M., Dalgarno, B., Gregor E., Kennedy, G.E., Lee, M.J.W. & Kenney, J. (2015). Design and implementation factors in blended synchronous learning environments: Outcomes from a cross-case analysis. *Computers & Education* Vol. 86: 1-17.
- Black, P. & William, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, Vol. 80.
- Brendborg, N. (2024). *Vanedyr*. København: Grønningen 1.
- Deslauriersa, L., McCartya, L.M., Miller, K., Callaghana, K. & Kestina, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *PNAS*.
- DfE (Department for Education) (2024). *Generative AI in Education. Educator and Expert Views*. Gov.uk. London: Department for Education.
- Elstad, E. (2023). Når læreren møter ChatGPT. Oslo: Universitetsforlaget.
- Ferguson, C., van den Broek, E.L. & van Oostendorp, H. (2022). AI-Induced guidance: Preserving the optimal Zone of Proximal Development. *Computers and Education Artificial Intelligence* 3.
- Flyverbom, M. et al. (2024). Grænser for techgiganternes udvikling og anvendelse af kunstig intelligens. *Delafrapportering 2 fra regeringens ekspertgruppe om techgiganter*. København: Erhvervsministeriet.
- Foug, S.S. (2023). Danske elevers læsekompetence i 4. klasse. Resultater af PIRLS-undersøgelsen 2021. Aarhus: Aarhus Universitetsforlag.
- Gadamer, H.-G. (2000). *Teoriens lovprisning – taler og artikler*. Aarhus: Systime.
- Haidt, J. (2024). Den ængstelige generation – hvordan vi vendte op og ned på barndommen og skabte en epidemi af mental sygdom. Aarhus: Svane & Bilgrav.
- Hattie, J.A.C. (2009). *Visible Learning. A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Abingdon: Routledge.
- Hattie, J.A.C. (2018). Ranking: 252 Influences and effect sizes related to student achievement. <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement>.
- Haug, A. & Haug, K.B. (2023). Sådan får vi mest ud af ChatGPT. *Politiken debat* den 10. april.
- Herløv, P. (2023). Lærere er bekymrede over elevers brug af chatbotter: Hvad må vi, og hvad må vi ikke? DR: Nyheder den 18. april 2023.
- Hunt, M.G., Marx, R., Lipson, C. & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, 37(10), 751-768.
- Håstein, H. (2017). *SMTTE: en struktur for pædagogisk opmærksomhed*. Tønsberg: Galleberg Forlag.
- Illeris, K. (2017). *Læringsteoriens elementer – hvordan hænger det hele sammen?* I: Illeris, K. (Red.). 49 tekster om læring. 4. oplag. København: Hans Reitzels Forlag.
- Kagan, S. & Stenlev, J. (2007). *Cooperative learning – undervisning med samarbejdsstrukturer*. København: Malling Beck.
- Kierkegaard, L., Pommerencke, L.M., Flensburg, S., Bonnesen, C.T., Madsen, K.R. (...) & Thorhauge, A.M. (2020). *Digital mediebrugs betydning for sociale relationer, fællesskaber og stress blandt børn og unge. En litteraturgennemgang*. København: Statens Institut for Folkesundhed.
- Koper, R. & Bennett, S. (2008). *Learning Design: Concepts*. I: Adelsberger, H.H., Pawlowski, J.M. & Sampson, D.G. (Eds.). *Handbook on Information Technologies for Education and Training*. International Handbooks on Information Systems. Berlin: Springer, 135-154.
- Labadze, L., Grigolia, M. & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education* 20, 56.
- Madsen, K.R., Román, J.E.I., Damsgaard, M.T., Holstein, B.E., Kristoffersen, M.J. (...) & Toftager, M. (2023). *Skolebørnsundersøgelsen 2022*. Hbsc.dk. København: Statens Institut for Folkesundhed, SDU.
- Mehlsen, C. (2024). *Opmærksomhedstyveriet*. København: Informations Forlag.
- Ottosen, M.H. & Andreasen, A.G. (2020). *Børn og unges trivsel og brug af digitale medier – to analysenotater*. København: VIVE.
- Pedersen, H.S., Flarup, L.H. & Arendt, K.S. (2018). *Faglig udvikling, trivsel og alsidig udvikling blandt elever med særlige behov. En redskabskortlægning og praksisafdækning*. København: VIVE.
- Pedersen, J., Rasmussen, M.G.B., Sørensen, S.O., Mortensen, S.R., Olesen, L.G. (...) & Grøntved, A. (2022). Effects of limiting digital screen use on well-being, mood, and biomarkers of stress in adults. *NPJ Mental Health Research* 1, 14.
- Pintrich, P.R. & Schunk, D.H. (2002). *Motivation in Education. Theory, Research, and Applications*. New Jersey: Prentice Hall.
- Rasmussen, M., Kierkegaard, L., Rosenwein, S.V., Holstein, B., Damsgaard, M.T. & Due, P. (red). (2019). *Skolebørnsundersøgelsen 2018. Helbred, trivsel og sundhedsadfærd blandt 11-, 13- og 15-årige skoleelever i Danmark*. København: Statens Institut for Folkesundhed.
- Rogers, E.M. (2003). *Diffusion of innovations*. New York, NY: Free Press.
- Sundhedsstyrelsen (2023). *Notat: Sundhedsstyrelsens anbefalinger om skærmb brug*. København: Sundhedsstyrelsen.
- Sørensen, N.U., Katznelson, N., Nielsen, M.L., Madsen, E.Aa. & Marstrand, B. (2024). *Levbar ungdomsliv: om unges bestræbelser på at trives*. København: CEFU Center for Ungdomsforskning.
- Webb, M.E., Prasse, D., Phillips, M., Kadjevich, D.M., Angeli, C. (...) & Laugesen, H. (2018). Challenges for IT Enabled Formative Assessment of Complex 21st Century Skills. *Technology, Knowledge and Learning*, 23(3), 441-456.
- Vembye, M.H. (2022). *Methodological Innovation in Systematic Reviewing and Statistical Meta-Analysis in Education and Beyond*. Ph.D. thesis. København: Danmarks institut for Pædagogik og Uddannelse (DPU), Aarhus Universitet.
- Winn, Z. (2023). *Study finds ChatGPT boosts worker productivity for some writing tasks*. MIT News July 14. Boston: Massachusetts Institute of Technology.