Sindre Mathias Strømnes Nordvoll

Automatiske tilbakemeldingssystemer i matematisk læring

Kvantitativ studie av automatisk gitte tilbakemeldingers effekt på prestasjoner

Masteroppgave i matematikkdidaktikk 5.–10. trinn
Veileder: Hermund André Torkildsen
Mai 2020

Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for lærerutdanning

NTNU
Kunnskap for en bedre verden

Sindre Mathias Strømnes Nordvoll

Automatiske tilbakemeldingssystemer i matematisk læring

Kvantitativ studie av automatisk gitte tilbakemeldingers effekt på prestasjoner

Masteroppgave i matematikkdidaktikk 5.–10. trinn
Veileder: Hermund André Torkildsen
Mai 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for lærerutdanning

NTNU
Norwegian University of
Science and Technology

Forord

Året og prosessen i å skrive en masteroppgave rangeres for meg sammen med utenlandsoppholdet i Afrika som det mest spennende og interessante i studietiden. Å få muligheten til å jobbe med et prosjekt i et slikt format og størrelse har vært pirrende. Jeg har vært heldig og kunnet vie ett helt år til å utfolde meg kreativt og faglig. Ikke bare har det oppvekket en videre interesse for datateknologi, programmering og matematisk forskning. En oppdagelse av arbeidsfeltet som utgjør pedagogisk og matematisk forskning, gjør at å komme tilbake til et forskermiljø ikke virker usannsynlig i fremtiden. Denne oppgaven hadde dog ikke vært den samme uten menneskene rundt meg.

Det er derfor jeg gjerne vil tillate meg å takke min kjæreste og vanligvis faste samboer Linnea for både faglig og emosjonell støtte gjennom året. Samtidig hadde jeg ikke fått samlet tankene mine uten hjelp fra både min veileder Hermund og datakunnskapene til Tore. I tillegg håper jeg at mine stadige bank på hans kontordør ikke har plaget Trygve altfor mye. Mine tre gode venner Håvard, Yohann og Kristian fortjener også en stor takk for faglig sparring, utfordring, latter og inspirasjon gjennom arbeidet.

Sammendrag

Denne oppgaven beskriver en metode for å la datamaskiner automatisk kunne gi tilbakemeldinger på ungdomsskoleelevers arbeid med matematikkoppgaver. I tillegg undersøkes det hvilken effekt ulike typer maskingitte tilbakemeldinger har på elevenes prestasjoner. Private aktører som Maple T.A., Kikora, Gyldendal og Khan Academy tilbyr nettoppgaver med ulike metoder for å automatisk evaluere elevsvar. De fleste av disse, og lignende tjenester, tilbyr en form for rett/gal verifisering av oppgavene.

Som motvekt til disse private leverandørene ble det utviklet en programvare og metode for å gi mer formative og kontekstuelt responderende tilbakemeldinger gjennom en datamaskin. Hvilke typer tilbakemeldinger og hvordan deres innhold genereres er inspirert etter Narciss & Huth (2002) og Narciss (2013) sin strategi under navnet feilrelatert veiledning. En slik strategi gir elevene assisterte og flerfoldige forsøk ved å gi strategisk nyttig informasjon for oppgaveløsning, fremfor å gi svaret direkte.

Programvaren som ble utviklet ble testet gjennom å la $n = 123$ tiendeklasseelever jobbe med en test i lineære funksjoner. Elevene skal i løpet av testen tilpasse en graf til ulike tabeller, funksjonsuttrykk og kontekstsituasjoner. Gjennom testen gir programvaren enten verifiserende eller formative og responsspesifikke tilbakemeldinger. Et semi-randomisert eksperiment med en kontrollgruppe som gis verifiserende tilbakemeldinger i testen benyttes for å avsløre hvordan de to typene tilbakemeldinger påvirker prestasjoner i testen.

Resultatene viser at den eksperimentelle gruppen som ble utsatt for formative maskingitte tilbakemeldinger, presterer moderat bedre målt i effektstørrelse ($g = 0.43$) ved $p = .023$. I tillegg brukte denne gruppen 57% lenger tid på testen. Samtidig er det store forskjeller på hvor mye tilbakemeldingene ser ut til å hjelpe elevene. På oppgaver som handler om å oversette en tabell til graf finnes det ingen signifikant forskjell mellom de to testgruppene i eksperimentet.

Samlet sett indikerer resultatene at datamaskiner kan ha en god effekt på prestasjoner dersom tilbakemeldingene de gir på matematikkoppgaver er utformet etter formative prinsipper. Likevel finnes det utfordringer knyttet til skalering av gode automatiske tilbakemeldingssystemer for et stort antall oppgaver. Metoden denne undersøkelsen er fundert på krever at tilbakemeldingene og tilstandene maskinen skal reagerer på, må konstrueres manuelt.

Innholdsfortegnelse

1 Innledning

Tilbakemeldinger ansees for å være et viktig konstrukt i mange læringsteorier. De kan gi nyttig informasjon til en elev om hvordan deres respons betraktes relativt til en endelig løsning, og hva som skal til for å komme til løsningen (Hattie & Timperley, 2007).

Fra behaviorismen kjenner vi tilbakemeldinger som forsterkning av ønsket atferd eller respons. Fra et slikt ståsted vil premiering av fullføring, fasit og selve løsningen kanskje vektlegges. Dermed vil informasjon om foreløpig prestasjon og verifisering av hvert enkelt svar prioriteres i tilbakemeldinger (Mory, 1996; Narciss, 2004). Dersom en betrakter tilbakemeldinger fra et kognitivt perspektiv kan tilbakemeldinger opptre som en nyttig kilde til nødvendig informasjon for å korrigere et feilaktig løst problem (Kulhavy & Stock, 1989). En utdypende type tilbakemelding vil generelt kunne gis som forklaring på hvorfor et svar er galt eller korrekt. Utdypende tilbakemeldinger kan gi nyttige opplysninger uten å avsløre den eksakte metoden eller det korrekte svaret før eleven får prøvd selv (Shute, 2008). De kan gis i form av hint, feilanalyser, ta opp systematiske feil, gi informasjon om temaet som adresses og mer (Narciss & Huth (2002).

Sosiokognitiv læringsteori fordrer at vi betrakter tilbakemeldinger fra et metakognitivt perspektiv. Det kan innebære å få tilbakemelding på ens egen atferd, deriblant strategier for oppgaveløsning, memoreringsteknikker, eller motspørsmål utformet for å trigge selvevaluering av atferd og tankemønster (Mevarech & Kramarski, 1997, 2003; Kramarski & Zeichner 2006).

Tidligere forskningsarbeid på skriftlige formative tilbakemeldinger viser dog at det ikke finnes noen bred enighet om hvilke konkrete typer tilbakemeldinger som gir best utslag på matematisk læring (Hattie & Timperley, 2007; Kulhavy & Stock, 1989). I enkelte tilfeller viser det seg også at tilbakemeldinger kan gi negativt utslag på prestasjoner og læring (Kluger & DeNisi, 1996). Til tross for dette blir enkelte typer skriftlige formative tilbakemeldingsmetoder fremhevet i flere artikler som relevante for å promotere læring (Narciss, 2013; Shute, 2008). I en slik sammenheng er det derfor viktig å forstå hvordan dette bilde kan nyanseres ved innføring av datamaskingitte tilbakemeldinger.

Det finnes midlertidig mange ulike private aktører som tilbyr ulike digitale plattformer som på ulikt vis evaluerer elevenes svar. Automatiske vurderingssystemer som Maple T.A. benyttes allerede i dagens utdanningsinstitusjoner i undervisningsøyemed. Maple T.A. har vært i bruk i flere år og er utbredt på mange universiteter i ulike land. Programvaren tar imot lekser og øvinger fra studentene, og vurderer hvorvidt svarene de oppgir er korrekte eller ei, før den gir en verifiserende beskjed som tilbakemelding. Det finnes erfaringsrapporter fra blant annet Italia (Barana et al., 2015; Barana & Marchisio, 2016), Norge (Rønning, 2017), Portugal (Brito et al., 2009), USA (Blackman, 2014) og Østerrike (Winkler et al., 2012) som hevder flere fordeler med bruk av et slikt system. De trekker frem tid og ressursbesparelser på korrigering, automatisk og umiddelbar tilbakemelding på studentbesvarelser, og i enkelte tilfeller, bedrede prestasjoner fra studentene.

Likevel fremstår aspektet med formativ tilbakemelding som en stor utfordring for oppgaver som vurderes av automatiske systemer (Morlandstø et al., 2019; Rønning, 2017). Tilbakemeldingene fra Maple T.A. leveres primært som simpel bekreftelse på svaret – altså en verifisering. Det er ønskelig at datamaskinen i større grad skal «forstå» meningsinnholdet i svaret en elev avgir. Den bør kunne agere kontekstuelt på responsen, og dette bør reflekteres i tilbakemeldingen. Shute (2008) rapporterer om effektiv tilbakemelding at «[...]*responsspesifikke tilbakemeldinger ser ut til å promotere elevers prestasjoner, spesielt læringseffektiviteten, mer enn andre typer feedback som simpel bekreftelse eller «svar til korrekt» vurdering*.» (s. 158).

Andre aktører som Khan Academy, Gyldendal Maximum Smartøving og Kikora leverer nettsider med matematisk innhold, der strømming av undervisningsvideoer, lesing av pensum og arbeid med oppgaver gjøres tilgjengelig nær sagt uavhengig av geografisk posisjon. Elever kan også interagere med koordinatsystemer og geometri gjennom grafiske brukergrensesnitt. Slike opplegg skaper en ny dimensjon for matematisk arbeid. Det svake punktet fremtrer også her som god tilbakemelding og evaluering av arbeidet elevene gjør på nett og i programvaresammenheng generelt. Oppgavene man finner i slike nettbaserte læringsressurser gir ofte mangelfull eller ingen tilbakemelding på utførte elevoppgaver i ulike matematiske emner. Akkurat som ved Maple T.A. er en av de mest utbrette typen tilbakemelding i denne sammenheng gjerne å stemple et svar med «gal» eller «riktig» - en ren verifisering. Eleven kan alternativt få selve svaret og løsningsstrategien opp ved feil besvarelse i Kikoras plattform. I tillegg presenteres elevene med virtuelle medaljer og fanfarer ved gjennomføring av en økt eller et sett med oppgaver.

I likhet med Shute (2008) skisserer Greenhow (2015) hvordan effektiv tilbakemelding bør administreres. Han påstår at vurdering ikke bare skal gi karakterbedømmelse og stimulere ytre motivasjon, men også promotere læring. Da er det helt essensielt å skrive tydelige og fullstendige tilbakemeldinger. Den store utfordringen alle de skisserte systemene har til felles er deres evne til å håndtere og adressere ulike typer ukorrekte eller mangelfulle svar. I praksis fører det til at den som konstruerer oppgavene også må forutse og planlegge for eventuelle feil eller misforståelser som elevene kan ta med seg til problemene.

I denne studien vil jeg derfor beskrive hvordan automatiske formative tilbakemeldinger gitt gjennom en datamaskin kan utformes, og hvilken effekt det har på et utvalg 10. klasseelever. 10. klasseelevene testes i digitale oppgaver om oversettelse av representasjoner av lineære funksjoner. Deretter sammenlignes deres prestasjoner med typen tilbakemelding de er gitt. Målet å tegne et bilde av hvordan automatiske tilbakemeldinger påvirker læring av funksjoner. Funksjoner er et grunnleggende tema som ligger nært knyttet til algebra, og har stort potensiale i digital sammenheng (Sacristan et al., 2010). Funksjonskompetanse er en essensiell del av matematikken, i nåværende og kommende læreplaner, samt et punkt hvor norske elever tradisjonelt gjør det dårlig ved TIMMS i internasjonal sammenligning (Grønmo & Hole, 2017).

1.1 Problemstilling

Undersøkelsen søker å finne svar på følgende spørsmål:

1. Hvordan kan det designes automatisk datamaskingitte, utdypende og verifiserende tilbakemeldinger til oppgaver om oversettelse fra tabellform, algebraisk- og situasjonsrepresentasjon til grafisk representasjon av lineære funksjoner?
2. Hvordan påvirker automatiske verifiserende og utdypende tilbakemeldinger gitt av en datamaskin elevers prestasjoner i oversettelsesoppgaver i lineære funksjoner?

Det første delspørsmålet vil bli besvart gjennom undersøkelsens metodiske kapittel, da det tar for seg designet av tilbakemeldingene og måten de leveres på. De to siste delspørsmålene settes opp mot hverandre gjennom et semi-randomisert eksperiment, og besvares følgende gjennom undersøkelsens resultat- og drøftingskapittel.

2 Teori

For å vurdere hvordan digitale automatiske tilbakemeldinger påvirker elevenes læring av og kompetanse i funksjoner må det omfattende begrepet funksjonskompetanse defineres for denne studiens øyemed. I første del av teorikapittelet undersøkes hvordan kompetanse i lineære funksjoner kan defineres i relasjon til representasjoner av funksjoner og overganger mellom dem. Deretter følger en beskrivelse av typiske feil og misoppfatninger som elever kan møte på i forbindelse med overganger til grafisk representasjon av lineære funksjoner.

Videre vil jeg diskutere hvordan representasjoner av funksjoner kan arbeides med i et digitalt medium. I delkapittelet etter undersøker jeg hvordan tilbakemeldinger kan gis av en digital programvare på en slik måte at det kan fremme elevenes prestasjoner, og nevnte kompetanse i overganger til grafiske representasjoner av lineære funksjoner. Til slutt vil noen av de større og mest brukte nettplattformenes tilbakemeldingsmodeller diskuteres for å sette min tiltenkte tilbakemeldingsmetode i sammenheng med nåværende løsninger.

2.1 Tilgang på matematiske objekter gjennom representasjoner

Denne undersøkelsen er fundert på ideen om at matematikk handler om abstraksjoner. I motsetning til observerbare fenomener i andre naturvitenskapelige forskningsgrener er ikke matematikk mulig å få tilgang til gjennom måleinstrumenter. Matematiske idéer, konsepter og sammenhenger – altså matematiske objekter, er tilgjengelig for bevisstheten via måten vi utviser dem på. Den franske semiotikeren Raymond Duval (2002; 2006) hevder at de matematiske objektene kun er tilgjengelige gjennom *representasjoner*, og at matematiske aktiviteter kun kan utøves og læres gjennom disse representasjonene. Han kaller representasjonene for semiotiske informasjonssystemer – systemer som benyttes i å kommunisere ulike matematiske idéer.

I tillegg benytter han begrepet representasjonsregister for å beskrive semiotiske systemer som kan *transformeres*, slik som det algebraiske, grafiske eller tabellformmessige registeret i forbindelse med funksjoner. Duval (2006) skiller mellom to typer transformasjoner: *behandling* og *oversettelse*. Behandling er transformasjoner som kan gjøres innad i ett representasjonsregister, for eksempel forenkling av funksjonsuttrykk. Oversettelse handler om transformasjoner mellom ulike type register uten å endre betydningen av objektet i konteksten, slik som å gå fra den algebraiske notasjonen for et objekt til dets grafiske representasjon (Duval, 2006). Duval advarer mot å forstå oversettelser som alltid ekvivalente i kompleksitet når en oversetter mellom representasjonsregister. Eksempelvis er det å gå fra tabellform til grafisk representasjon en annerledes behandling enn å gå motsatt vei. Et skille mellom retningen i oversettelser, altså hva som defineres som kilderepresentasjonen og målrepresentasjonen oppstår (Duval, 2006; Bossé, Adu-Gyamfi, & Cheetham, 2011). Kilderepresentasjonen er representasjonsregisteret en transformerer fra og målrepresentasjonen vil da defineres som registeret som det transformeres til.

Videre i denne undersøkelsen vil begrepet representasjon referere til et konvensjonelt informasjonssystem som kan utvise enkelte, men ikke nødvendigvis all informasjon om objektet som representeres (Adu-Gyamfi & Bossé, 2014).

2.1.1 Funksjonskompetanse

For at en datamaskin skal kunne gi tilbakemeldinger på elevers arbeid med funksjoner, bør kompetanse innenfor tema funksjoner defineres. Forskning på funksjonskompetanse belyser ulike delelementer som inngår i funksjonskompetanse.

Brian O'Callaghan (1998) oppsummerer de ulike synene på funksjonskompetanse. Han beskriver de ulike kompetanseområdene under funksjonsbegrepet som *modellering*, *tolkning*, *oversettelse*, *tinglinggjøring* (reifying) og et akkompagnert sett med *prosedyreferdigheter* som er underliggende for å mestre de fire øvrige kompetanseområdene av funksjoner. Modelleringskompetanse relateres til problemløsing, og defineres som en persons evne til å representere en problemsituasjon eller situasjonsbeskrivelse gjennom bruk av funksjoner (O'Callaghan, 1998). O'Callaghan hevder at modellering av situasjoner for å organisere informasjon fra den virkelige verden er en av de mest vanlige bruksområdene til funksjoner generelt.

Når det gjelder tolkning beskriver O'Callaghan (1998) dette kompetanseområdet som den omvendte prosessen av modellering. Tolkning handler derfor om å forstå ulike situasjonsbeskrivende representasjoner fra den virkelige verden, for eksempel ved å forstå innholdet i en graf fra en avis. Tolkning vil også handle om å forstå delelementer i en funksjonsrepresentasjon lokalt (de individuelle punktene i en graf, et uttrykks stigningstall etc.) og globalt (graf, tabell eller funksjonsuttrykk som helhet) (O'Callaghan, 1998).

O’Callaghans (1998) tredje kompetanseområde for funksjoner beskrives som oversetting. Synonymt med det Duval (2006) beskriver som oversettelse, er det ens evne til å utføre overganger mellom de ulike representasonene til en funksjon. Dette tilfellet kan blant annet dreie seg om å oversette et algebraisk funksjonsuttrykk eller en numerisk tabell til grafisk representasjon.

Tingliggjøring beskrives som det fjerde og siste isolerte kompetanseområdet for funksjoner. Det defineres som skapelsen av et mentalt, matematisk objekt av det som tidligere ble ansett for å være en prosess eller prosedyre (O'Callaghan, 1998; Sfard, 1991). Tingliggjøring ansees som det endelige nivået i funksjonsforståelse og blir derfor ikke nådd av mange elever, men beskrives samtidig som essenesielt for å kunne gjøre mer komplekse operasjoner med funksjoner.

Som nevnt tidligere forbindes disse fire kompetanseområdene med et sett prosedyrekunnskaper (O'Callaghan, 1998). Prosedyrekunnskapene er elementet som gjør det mulig for elevene å arbeide med funksjoner gjennom representasjonssystemene. Det kan blant annet omhandle relevante algebrakunnskaper, forståelse av hvordan et kartesisk koordinatsystem kan brukes eller kunnskaper om numeriske tabellers oppbygging (O'Callaghan, 1998).

I tillegg vil jeg påpeke at det finnes et siste kompetanseelement som er svært relevant for læring av funksjoner – begrepeskompetanse. Zalman Usiskin er kanskje mest kjent for sitt arbeid med Van Hiele geometritester, og er kanskje derfor generelt opptatt av å fremheve matematikkens vokabular. Vi har ord for omtrent samtlige matematiske symboler og konsepter, og kjennskap til det matematiske språket er forgjengeren til all videre forståelse hevder han (Usiskin, 2012). Forståelse av begreper som *stigningstall*, *konstantledd* og *skjæringspunkt med y-aksen* kan med andre ord sies å være svært relevante for elevers oppbygging av funksjonskompetanse.

De fleste studier er likhet med O’Callaghan opptatt av å undersøke forståelse av funksjoner som tolkning av og oversettelse mellom representasjoner. Leinhardt, Zaslavski & Stein (1990) påstår at hovedvekten av funksjonsoppgaver i ulike studier baserer seg på to kognitive handlinger. De beskrives som *tolkning* (avlesing og innhenting av informasjon) og *konstruksjon* (plotting av graf, konstruere uttrykket til en graf, eller generering av eksempel på funksjon). Variasjoner innen disse to handlingene skaper i følge forskerne de ulike typene oppgavene som finnes innen arbeid med funksjoner. De utdyper videre at alle oppgavetypene kan videre varieres ved å innføre *situasjoner* – altså en kontekst som funksjonen opererer innenfor (Leinhardt et al., 1990).

Nyere forskning støtter opp under tankegangen til Leinhardt et al. (1990), og presiserer typisk at det å kunne oversette mellom ulike representasjonsformer er kritisk for å forstå funksjoner. (Ainsworth, Bibby, & Wood, 2002). En slik oversettelse brukes i enkelte studier synonymt med *translasjon* eller *omdanning*, og defineres som den kognitive prosessen i å transformere informasjonen iboende i et matematisk objekt. Oversettelsesoppgaver vil typisk handle om å gjenkjenne den samme funksjonen i ulike representasjonsformer, eller å konstruere en representasjon gitt en annen. Altså vil oversettelsesoppgaver handle om *enten* tolkning eller konstruksjon av representasjoner.

> By translation we refer mainly to (a) the act of recognizing the same function in different forms of representations; (b) identifying for a specific transformation of a function in one representation its corresponding transformation in another representation (Kaput, 1987b, 1987c; Yerushalmy, 1989); or (c) constructing one representation of a function given another one. (Leinhardt et al., 1990, s. 16).

Leinhardt et al. (1990) eksemplifiserer utsagnet over med en oppgave der elever blir bedt om å tolke flere funksjonsuttrykk opp mot deres korresponderende grafiske representasjon. En slik oppgave krever at eleven kan bevege seg smidig mellom de algebraiske funksjonsuttrykkene og grafene for å tolke deres meningsinnhold.

Forskerteamet bestående av Nitsch, Fredebohm, Bruder, Kelava, Naccarella, Leuders & Wirtz (2015) beskriver i likhet med Leinhardt et al. (1990) hvordan oversettelse mellom ulike represenstasjonformer av funksjoner innebærer å utøve ulike kognitive handlinger. Oversettelsesoppgaver blir beskrevet som viktige arbeidsverktøy for å skaffe seg innsikt i elevenes forståelse av funksjoner fordi denne typen oppgaver krever to spesielle handlinger av elevene. De konkluderer i sin analyse av kompetansemåling i funksjoner at «[...]de to grunnleggende handlingene identifisering og konstruksjon er nyttige for å identifisere elevers forståelse av konseptet funksjoner.» (Nitsch, et al., 2015, s. 676). Deres tredelte modell beskriver *identifisering*, *konstruksjon*, samt *beskrivelser/forklaringer* som disse kognitive handlingene.

Det å *identifisere* spesielle verdier eller egenskaper i en representasjon fremtrer som grunnleggende essensielt for alle de resterende handlingene, mens konstruksjon skiller seg ut ved at eleven må selv skape målrepresentasjonen fra en kilderepresentasjon (Nitsch, et al., 2015). Skal du oversette en funksjon i form av et algebraisk uttrykk til en graf må en altså kunne identifisere uttrykkets egenskaper før grafen kan konstrueres. Den siste kognitive handlingen, beskrivelser/forklaringer, er også unike ved at de krever gode kommunikative evner av elevene. De må inne ha god kunnskap om oppgaven de har jobbet med og være bevisst deres egen løsningsprosess for å kunne beskrive den. Det vil være verdt å merke seg en vesensforskjell mellom begrepene man finner i ulike studier. Leinhardt et al. (1990) skiller mellom *tolkning* og *konstruksjon*, der Nitsch et al. (2015) deler mellom *identifisering* og *konstruksjon*. Forskjellen ligger i at begrepet tolkning ikke nødvendigvis er knyttet direkte til oversettelsesprosessen, men heller handler om å forstå innholdet i deler eller helheten i en representasjonen. Identifisering betegner handlingen i å identifisere bestemte egenskaper eller verdier i en funksjon, som dets stigningstall, ekstremalpunkter eller skjæringspunkt med y-aksen.

2.1.3 Læreplanens kompetansemål for funksjoner

LK06 regulerer hva kompetanse i funksjoner betyr for norske elever. Læreplanen inneholder følgende kompetansemål for 10. trinn i funksjoner:
«Mål for opplæringen er at eleven skal kunne

- lage funksjonar som beskriv numeriske samanhengar og praktiske situasjonar, med og utan digitale verktøy, beskrive og tolke dei og omsetje mellom ulike representasjonar av funksjonar, som grafar, tabellar, formlar og tekstar
- identifisere og utnytte eigenskapane til proporsjonale, omvendt proporsjonale, lineære og kvadratiske funksjonar og gje døme på praktiske situasjonar som kan beskrivast med desse funksjonane» (Utdanningsdirektoratet, 2014).

Vi ser altså hvordan læreplanen sier seg enig i både Leinhardt et al. (1990) og Nitsch et al. (2015) sin vurdering av hva funksjonskompetanse er. Det første kulepunktet kan sies å handle om den kognitive handlingen konstruksjon, mens det andre beskriver identifiseringshandlinger. I tillegg gir LK06 rom for å arbeide med oversetting mellom ulike funksjonsrepresentasjoner med digitale verktøy. Det nevnes også at bl.a. lineære funksjoners egenskaper bør identifiseres og utnyttes, samt ulike praktiske situasjoner som kan beskrives med slike funksjoner.

Den nye læreplanen gir også en pekepinn på hva som bli viktige overordnede kompetansemål for 10. årstrinn i funksjoner for perioden som kommer fra og med 2021:
«Mål for opplæringa er at eleven skal kunne

- utforske og samanlikne eigenskapar ved ulike funksjonar ved å bruke digitale verktøy
- rekne ut stigingstalet til ein lineær funksjon og bruke det til å forklare omgrepa endring per eining og gjennomsnittsfart
- utforske samanhengen mellom konstant prosentvis endring, vekstfaktor og eksponentialfunksjonar
- bruke funksjonar i modellering og argumentere for framgangsmåtar og resultat» (Utdanningsdirektoratet, 2020).

Her ser vi at ulike funksjoners egenskaper nå skal utforskes og sammenlignes gjennom bruk av digitale verktøy. I tillegg ser vi at noen av disse egenskapene, som stigningstallet til en lineær funksjon, helt konkret er nedfelt i den nye læreplanen. Bruk av modellering i sammenheng med funksjoner nevnes også. Gitt at modellering kan defineres som prosessen i å oversette mellom den virkelige verden og matematikk, kan det argumenteres for at oversettelse fra situasjonsbeskrivelser til grafisk representasjon er relevant (Blum & Ferri, 2009).

Det å forstå de ulike representasjonsformene for funksjoner er altså en svært viktig del av funksjonskompetanse, spesielt tatt i betraktning av hvordan tilgang på matematiske objekter som funksjoner er begrenset til nettopp representasjonene vi har for dem (Duval, 2006). Denne undersøkelsen søker å finne svar på hvordan elevers feil i arbeid med funksjons kan oppdages av en programvare, for å videre kunne gi automatisk tilbakemelding og spore dets effekt. Derfor vil funksjonskompetanse videre i teksten brukes som å inneha gode oversettelsesferdigheter mellom ulike representasjonsregistre – mer spesifikt at en evner å kunne identifisere egenskapene til kilderepresentasjoner for å konstruere målrepresentasjoner av funksjoner. I tillegg vil forståelse av en lineær funksjons stigningstall og skjæringspunkt med y-aksen vektlegges som viktige elementer i denne prosessen (Chiu, Kessel, Moschkovic, & Muñoz-Nuñez, 2001). Videre er det nødvendig å avgjøre hvilke representasjonsregistre som er egnede for å automatisk bedømme hvilke feil som er gjort, og hvilken tilbakemelding som er høvelig i møte med oppgaver om lineære funksjoner.

2.1.4 Representasjonsformer av lineære funksjoner

Noen studier hevder det i hovedsak finnes tre hovedrepresentasjoner; grafer, tabeller og likninger/uttrykk, hvorpå situasjonsbeskrivelser fremtrer som en måte å kontekstualisere eller innramme representasjonen på (Kaput 1992; O'Callaghan, 1998). I sistnevnte studier blir det å oversette en problemsituasjon til en matematisk representasjon som likning, tabell eller graf betegnet som *modellering*. Oversettelser fra den virkelige, ikke-matematiske verden til en matematisk funksjon regnes altså som en separat kompetanse i å forstå funksjonsbegrepet. Oversettelse mellom representasjoner er i disse studiene er derfor begrenset til overganger mellom grafer, tabeller og likninger.

Andre undersøkelser går ut ifra at det finnes fire hovedrepresentasjoner av funksjoner – grafer (G), numeriske tabeller (N), algebraiske likninger (A) og situasjonsbeskrivelser (S) (Nitsch et al, 2015; Adu-Gyamfi, Bossée & Stiff, 2011; Sierpinska, 1992; Duval, 2002). Her beskrives alle overganger mellom ulike representasjoner, matematiske eller ei, som en oversettelsesprosess. Argumentet for denne tolkningen ligger i at det ikke er representasjonen som oversettes, men heller idéene som uttrykkes gjennom dem. Denne undersøkelsen tar i likhet med Nitsch et al. (2015) videre utgangspunkt i at å konstruere overganger til grafer (G) fra numeriske tabeller (N), algebraiske likninger (A) og situasjonsbeskrivelser (S) kan regnes som oversettelsesoppgaver. Det neste steget blir dermed å beskrive hva som påvirker vanskegraden i forskjellige typer oversettelsesoppgaver til grafisk representasjon. Å ha kjennskap til ulike vanskegrad-faktorer vil gjøre det mulig å kontrollere hvilke elementer som kan være utfordrende for elever i testen jeg gjennomfører.

2.1.5 Overganger mellom grafiske og ikke-grafiske representasjoner – hva påvirker deres vanskegrad?

Forskning på funksjonskompetanse og oversettelseskompetanse fokuserer også på elevers vansker relatert til tolkning og oversettelse av funksjoner (Ainsworth, Bibby, & Wood, 2002; Adu-Gyamfi & Bossé, 2014; Bossé, Adu-Gyamfi, & Cheetham, 2011; Nitsch, et al., 2015).

Studier som inkluderer oversettelsesoppgaver vektlegger ulike typer representasjoner av funksjoner, men Leinhardt et al. beskriver hvordan de fleste lærebøker fokuserer på sammenhenger mellom grafiske og algebraiske representasjoner (1990, s.16). Spesielt ofte fremtrer oppgaver der eleven gjennom den kognitive handlingen tolkning blir bedt om å beskrive visse egenskaper til en funksjon, som å betrakte en graf, og deretter bestemme dets stigningstall. Færre tar for seg oppgaver som krever at elevene må selv konstruere en målrepresentasjon. Det er ønskelig at elever kan oversette på tvers av alle representasjonsformene, men i denne masteroppgaven undersøkes dette gjennom oversetting fra tabeller, uttrykk og situasjonsbeskrivelser til graf.

De ulike oversettelsesoppgavene det gir rom for å jobbe med vil kunne ha ulik vanskegrad. Vanskegraden i oversettelsesoppgaver kommer av ulike faktorer som er iboende i selve tolkningshandlingen som kreves for å gjennomføre oversettelsen (Janvier, 1987; Leinhardt et al., 1990; Nitsch et al., 2015). Samtidig vil kompleksiteten i en oversettelse ikke bare bestemmes ut ifra handlingen eleven må utføre i oversettelsesprosessen, men også ut ifra flere egenskaper ved kilde- og målrepresentasjonen, som *faktagap*, *forvirrende fakta* og *attributt-tetthet* (Bossé et al., 2011). Faktagap er avhengig av oppgavens mål og kontekst, da det handler om hvor mye informasjon som er utelatt fra en representasjon relativt til hva oppgaven etterspør. Om en numerisk tabell ikke inneholder skjæringspunktet med y-aksen (dvs. punktet $(0, f(0))$), kan tabellen sies å ha noe høyere faktagap dersom oppgaven etterspør dette (Adu-Gyamfi et al, 2012; Bossé et al., 2011).

Forvirrende fakta handler derimot om hvor mye «unødvendig» eller tilgjort informasjon en representasjon inneholder. En numerisk tabell kan for eksempel settes opp i en ikke-sekvensiell rekkefølge med tanke på x, eller ha flere x-verdier som assosieres med den samme y-verdien for å øke mengden forvirrende fakta ved representasjonen (Adu-Gyamfi et al, 2012; Bossé et al., 2011).

Til slutt definerer Adu-Gyamfi et al. (2012) attributt-tetthet som hvor mye relevant informasjon en representasjon inneholder, og hvor mye innsats som kreves for å avdekke øvrige egenskaper. Lav attributt-tetthet i en representasjon vil derfor bety at det kreves betydelig arbeid for å avdekke funksjonens kritiske egenskaper, som stigningstall og konstantledd. Senket attributt-tetthet opptrer som en av de vanligste metodene for å øke kompleksiteten i en oppgavetype (Bossé et al., 2011). Som eksempel skal en tenkt elev oversette et algebraisk uttrykk til grafisk representasjon. Oppgaven inneholder flere ledd enn det sedvanlige $y = ax + b$ – la oss si $y = 4x + 2 - 4 + 2x$. Den økte mengden ledd krever at eleven transformerer uttrykket til $y = 6x - 2$ for å kunne oversette enklere. Flere handlinger innad i en oversettelsesprosess vil føre til mindre sannsynlighet for å lykkes med oversettelseen (Adu-Gyamfi et al., 2012).

I de tilfeller det gjelder å konstruere en målrepresentasjon, vil representasjoner som opptrer som kilde og mål bestemme hva slags oversettelseshandling elevene må utøve slik det vises i tabell 1 under.

To / From	Situations, Verbal Description	Table	Graph	Formulae [Symbolic]
Situations, Verbal Description		Measuring	Sketching	Modeling
Table	Reading		Plotting	Fitting
Graph	Interpretation	Reading off		Curve fitting
Formulae [Symbolic]	Parameter Recognition	Computing	Sketching	

Tabell 1, oversettelseshandlinger (Bossé et al., 2011, s. 119)

Gjennom å oversette fra en numerisk tabell til grafisk representasjon (T→G) kreves det at eleven kan *plotte* punktene i tabellen i et koordinatsystem. Plotting er en handling som ansees for å være *lokal* (Adu-Gyamfi et al., 2012; Bossé et al., 2011; Leinhardt et al., 1990). Begrepet lokal handling henviser til at eleven har en punktvis tilnærming til representasjonen for hånden. Når en leser av de ordnede parene i en numerisk tabell kreves det kun at koordinatene overføres korrekt til koordinatsystemet. Lokale handlinger regnes for å være enkle oversettelseshandlinger da hele funksjonen ikke må betraktes. Tabeller som kilderepresentasjon har vanligvis lav attributt-tetthet da de inneholder begrenset informasjon om funksjonens kritiske egenskaper som dets stigningstall uten å gjøre beregninger. I tillegg karakteriseres de generelt av lavt faktagap og lite forvirrende fakta, men disse faktorene kan modifiseres i et forsøk på å øke eller senke kompleksiteten i en oversettelse.

Elever kan også jobbe med å oversette fra algebraisk representasjon til graf, (A→G). En slik oversettelse krever at eleven kan *skissere* en graf, en handling som beskrives som global (Adu-Gyamfi, 2012; Bossé et al., 2011). Globale oversettelseshandlinger regnes som betydelig mer komplekse enn lokale, da de krever at oversetteren holder oversikt over den simultane variasjonen mellom to variabler i f.eks. et lineært uttrykk. I tillegg innehar algebraiske likninger høy attributt-tetthet som følge av at mange kritiske egenskaper fremkommer direkte i uttrykket. Punktene som inngår i grafen må selv bestemmes, men endringen mellom punktene og skjæringspunktet med y-aksen vises eksplisitt. Sammen med lite forvirrende fakta og lavt faktagap konkluderer forskerne med at denne oversettelsen er noe mer kompleks enn å oversette fra tabell til graf, da skissering regnes som en global handling.

Arbeid med å oversette fra situasjonsbeskrivelse til graf, (S→G) minner mye om A→G. Oppgaver som forespør en slik oversettelse innebærer i likhet med A→G at eleven kan *skissere* en graf, en global handling. Samtidig kan kontekstuelle situasjoner ha både lav og høy attributt-tetthet avhengig av informasjonen de gir elevene, noe som gir stort rom for variasjon i vanskegrad på dette punktet. Bossé et al. (2011) påpeker at S→G oversettelser generelt vil ha høyere attributt-tetthet enn numeriske tabeller, men lavere enn algebraiske uttrykk som kilderepresentasjon. Slike oppgaver er skrevet med et annet, mer talelikt språk enn det formelle symbolspråket matematikken ofte representeres gjennom. Dette gjør at atributt-tettheten til situasjonsbreskrivende oppgaver ofte kan være noe mindre enn ved algebraiske uttrykk, som gjerne er kompakte og inneholder mye informasjon om funksjonen som helhet.

På den annen side åpner den tekstuelle faktoren ved slike oppgaver for mange forvirrende fakta og høyt faktagap. Den virkelige verden og oppgaver som forsøker å beskrive sider ved den, kan uttrykke de mer rotete og ikke alltid så entydige aspektene en finner i de erfaringslevde, hverdagslige fenomener (Doorman & Gravemeijer, 2009). Verden utenfor det rent matematiske domenet lar seg ikke alltid kategorisere så enkelt som en kanskje skulle ønske – fenomener og variabler som tilsynelatende kan påvirke et problem kan være avhengig av andre betingelser enn først antatt. Det kan være vanskelig å bestemme hvilke fakta i en situasjonsbeskrivende oppgave som er relevante for å konstruere dets grafiske representasjon. Oppgaver som krever S→G- oversettelse kan ofte inkludere variabler eller tall som er irrelevante eller uviktige for å løse oppgaven, og på denne måten øke mengden forvirrende fakta (Bossé et al., 2011). Motsatt kan de også øke mengden faktagap ved gjøre essensiell informasjon for å løse oppgaven mindre synlige, ved at informasjon i teksten må dekodes og deretter beregnes, for å finne opplysningene som trengs for å konstruere grafen. Oppsummert sett betyr det at slike oppgaver vil kunne klassifiseres som å være de mest komplekse av de som er skissert.

2.2 Konstruksjon av grafiske representasjoner – elevers misoppfatninger og feil

I arbeid med å oversette til og konstruere ulike representasjoner av funksjoner finnes det en rekke typiske feil elever kan gjøre. Mange av disse feilene henger ofte sammen med flere faktorer som gjør oversettelsesoppgaver mer eller mindre vanskelige. Derfor er undersøkelser på vansker med oversettelse mellom representasjoner typisk knyttet til de ulike representasjonenes natur eller elevens aktivitet i arbeidet (Bossé et al., 2011; Adu-Gyamfi et al., 2012). Som skissert tidligere vil enkelte oversettelser av representasjoner kreve ulike oversettelsesteknikker, noe som kan føre til varierende vanskegrad. I tillegg kan noen representasjonsoppgaver være mer komplekse gjennom at de krever flere steg i oversettelsesprosessen. Elevene seg imellom kan også bruke ulike teknikker når de oversetter mellom representasjoner, og noen av disse kan være mer kompliserte enn andre (Janvier, 1987). Noen lærere kan også vektlegge ulike representasjoner og oppgaver mer enn andre. Det vil i disse tilfellene være mer naturlig at elevene i større grad mestrer den typen representasjoner de er mest kjent med, noe som vil variere mellom klasser og skoler (Bossé et al. 2011; Nitsch et al., 2015).

På et generelt nivå vil elever ofte utvise flere typer feil når de oversetter mellom grafiske- og algebraiske representasjoner og numeriske tabeller. Litteratursammenfatningen av Bossé, Adu-Gyamfi & Cheetham (2011) peker på *manipulasjonsfeil* og *konseptuelle feil* som typiske for elever. Manipulasjonsfeil beskrives som feil der eleven beregner et aritmetisk eller algebraisk problem på gal måte eller misforstår bruken av ulike variabler. Denne typen feil karakteriseres av misforståelser av stigningstallet til en funksjon og dets rolle, forvirring rundt intervall og punkter, andre nøkkelkonsepter som kreves for å konstruere målrepresentasjoner, eller generelle utregningsfeil (Bossé et al, 2011).

Konseptuelle feil oppstår når eleven misforstår oppgavens underliggende konsepter, eller har brukt ukorrekt logikk for å komme frem til svaret (Bossé et al., 2011). Når elever gjør konseptuelle er det mulig at utregninger som er gjort er korrekte. Dersom de har misforstått et essensielt konsept og dermed brukt feil metode for å representere eller beregne svaret, kan de ha utført hvert steg nøyaktig, men fortsatt ende opp med feil svar. Konseptuelle feil kan også handle om å overse en kritisk begrensning ved en oppgave, som når en funksjon beskriver en diskret situasjon, men behandles som om den var kontinuerlig – eksempelvis beregning av antall enheter et figurtall er bygget opp av. Konseptuelle feil preges også av syntaktiske misoppfatninger, som feil i tolkning av numeriske tabeller eller algebraiske uttrykk.

En studie av Adu-Gyamfi, Bossé , & Stiff (2012) oppsummerer på den annen side de ulike elevfeilene forbundet med oversettelsesoppgaver mellom grafer, numeriske tabeller og algebraiske uttrykk i tre hovedkategorier: *implementeringsfeil*, *tolkningsfeil* og *bevaringsfeil*. I tillegg beskriver de noen typiske feil i oversettelsesoppgaver fra situasjonsbeskrivelser til grafisk representasjon.

2.2.1 Implementeringsfeil

Implementeringsfeil oppstår ofte når et steg i en algoritme er utført på gal måte – altså at eleven har gjort en «utregningsfeil». Implementeringsfeil kan også komme til syne hvis en elev for eksempel endrer betydningen av $(-2, 1)$, og tolker -2 som y-verdien og 1 som x-verdien, eller ikke tar med fortegn i oversettelsen. Generelt vil algebraiske representasjoner ha høy attributt-tetthet ettersom de inneholder lite unødvendig informasjon som må behandles, men likevel kan nettopp denne faktoren gjøre at elever får implementeringsfeil på A→G oversettelsesoppgaver. Enkelte elever vil kanskje måtte beregne noen koordinater gjennom en tabell ut ifra likningen for å kunne plotte grafen. Det økte antallet steg i oversettelsen kan dermed gi større sannsynlighet for å få implementeringsfeil underveis. I oversettelse direkte fra numerisk tabell til graf vil altså implementeringsfeil ikke oppstå ettersom eleven ikke behøver å utføre en operasjon eller algoritme når hun skal plotte punkter (Adu-Gyamfi & Bossé, 2014; Adu-Gyamfi et al., 2012).

2.2.2 Tolkningsfeil

Den andre typen feil som beskrives i studiene, tolkningsfeil, skisseres som når en elev ukorrekt karakteriserer attributter eller egenskaper til enten kilde- eller målrepresentasjonen. Adu-Gyamfi et al. (2012) påpeker at en vanlig beskrivelse av en slik feil er at «eleven forstår ikke hvordan hun skal tolke eller bruke en egenskap» som er definerende for kilde- eller målrepresentasjonen (2012, s. 163). Eksempelvis kan elev med denne typen feil lese det algebraiske uttrykket $y = 3x + 4$, og tolke det som at funksjonens grafiske representasjon bør passere gjennom punktene $(3, 0)$ og $(0, 4)$. Det å blande mellom stigningstall og konstantledd i oversettelse fra algebraisk uttrykk til graf kan derfor også forstås som en tolkningsfeil ettersom eleven ikke klarer å tolke de kritiske egenskapene iboende i en lineær funksjon.

Tolkningsfeil kan også dukke opp i forbindelse med oversetting fra numerisk tabell til grafisk eller algebraisk representasjon, til tross for at spesielt oversetting til graf fra tabell beskrives av flere undersøkelser som relativt enkel prosess (Adu-Gyamfi et al., 2012; Leinhardt et al., 1990; Nitsch et al., 2015). Eleven trenger kun å vite at i en mengde med ordnede par vil hver av disse ordnede parene assosieres med et punkt i koordinatsystemet. Så lenge en kjenner til dette prinsippet trenger eleven ikke nødvendigvis forstå eller vite noe om en funksjons kritiske egenskaper for å plotte dets punkter. Likevel kan elever som evner å plotte punkter fra tabeller fortsatt gjøre feil som følge av senket attributt-tetthet, økt mengde faktagap eller forvirrende fakta i kilderepresentasjonen. Dersom en tabell som representerer en lineær funksjon f.eks. ikke lister opp skjæringspunktet med y-aksen, benytter desimaler, negative, svært små eller store tall, eller inneholder svært mange ordnede par vil det øke sannsynligheten for at elevene gjør flere feil. I tillegg kan bl.a. funksjonens stigningstall maskeres i tabellen gjennom at ulike verdier ikke listes opp i stigende rekkefølge eller ved å liste opp ikke-uniforme inkrementeringer av x-verdier (Bossé et al., 2011).

2.2.3 Bevaringsfeil

Til sist forklarer Adu-Gyamfi et al. (2012) den siste typiske elevfeilen – bevaringsfeil. Denne typen feil karakteriseres som at eleven har bevart likhet mellom kilde- og målrepresentasjon for egenskapene hun selv har identifisert, men mangler andre relevante egenskaper. Artikkelforfatterne viser til at denne typen feil i all hovedsak oppstår når elever skal oversette fra numerisk tabell eller algebraisk uttrykk til graf. Om en elev har kalkulert noen x- og y-koordinater fra et algebraisk uttrykk for å så plotte disse punktene, vil grafen ofte være korrekt representert for de kalkulerte punktene, men ikke nødvendigvis for øvrige punkter i grafen. Feilen ligger i dette tilfellet i at eleven på galt vis har forlenget et linjesegment.

2.2.4 Elevfeil i oversettelse fra situasjonsbeskrivelser til graf

Situasjonsbeskrivelser eller kontekstuelle oppgaver skiller seg ut fra de øvrige matematiske representasjonene ved at de er rent ordbaserte. Oppgaver som har situasjonsbeskrivelser som kilderepresentasjon har likevel to typiske feiltrekk (Bossé et al., 2011). Den første av slike feil blir beskrevet som ordrekkefølge-matchingfeil, og opptrer når eleven simpelthen antar at rekkefølgen på enkelte nøkkelbegreper i oppgaveteksten skal passe med rekkefølgen av variabler i likningen som beskriver den. Dette fenomenet er kjent i mange oppgavekontekster også utenfor funksjoner og likninger under begrepet *syntaktisk strategi*, men kan altså her fremstå som at eleven tror det første nøkkelbegrepet kan passe funksjonens stigningstall og det neste kan definere dets konstantledd (Clement, Lockhead, & Monk, 1981). Den siste typiske feilen i oversettelse fra situasjonsbeskrivelser er i følge Bossé et al. (2011) den statiske sammenligningsprosessen, der eleven forstår meningsinnholdet i oppgaveteksten, men vet ikke hvordan hun skal representere det videre. Den siste feiltypen minner altså om en form for konseptuell feil.

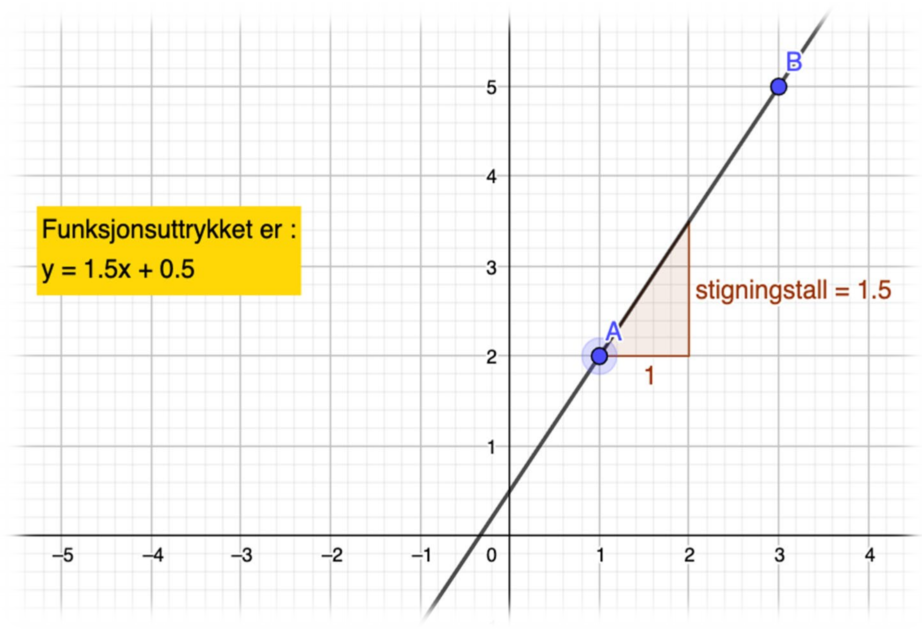
Bevaringsfeil, tolkningsfeil, syntaktisk strategi og statisk sammenligningsprosess fremstår som nyttige for denne undersøkelsen sin del. De lar meg peke på konkrete deler av oversettelsesprosessen til grafisk representasjon som kan gå galt. Feiltypene gjør det mulig å peke på spesielle situasjoner der elevene har vansker, for å deretter gi dem veiledende og formative tilbakemeldinger. Videre bør det avklares hvordan arbeid med funksjoner og oversettelser mellom dets representasjoner endres når operasjonene gjøres i et digitalt medium.

2.3 Representasjoner, digitale verktøy og dets egenskaper

Digitale teknologier og verktøy kan påvirke innlæring av matematikk på ulike vis, slik som muligheten til å fremvise *flere representasjoner* samtidig (Sacristán, et al., 2010). Videre innehar digitale verktøy flere *dynamiske* egenskaper, men kan også bidra med stor *beregningskapasitet* og umiddelbar *respons* til elevene (Ainley, 2001; Bueie, 2011; Kaput, 1992; Sacristán, et al., 2010).

Et sentralt aspekt ved digitale verktøys bidrag til matematikkundervisningen er altså deres evne til å vise flere representasjoner på samme tid. Sacristan et al. (2010) påpeker at bruken av flerfoldige representasjoner er gunstig, fordi en dynamisk matematikkprogramvare uttrykker ulike aspekter ved en funksjon tydeligere. Altså vil opplysninger elevene får fra å kombinere representasjoner være større enn ved enkeltrepresentasjoner.

En graf som settes opp kan synkront vise dets funksjonsuttrykk, stigningstall, konstantledd, toppunkt og mer. Det generelle uttrykket forblir det samme, men variablene kan dynamisk endre seg, slik at deres essensielle egenskaper kommer tydeligere frem (Sacristán, et al., 2010). Elevene kan altså få muligheten til å «oppdage» hva de ulike variablene i et algebraisk funksjonsuttrykk har å si for den grafiske representasjonens visuelle utseende, på en måte som ikke ville vært like enkelt å fremvise i statiske medier. Figur 1 under viser et eksempel på et enkelt skript som dynamisk viser et uttrykk og en graf samtidig.



Figur 1, dynamisk visning av to representasjoner, (egen produksjon)

I typisk statiske medium som penn og papir vil ikke en funksjon eller graf over tid kunne endre seg etter den er konstruert (Bueie, 2011; Sacristán, et al., 2010). Med mindre eleven visker ut representasjonen eller lager en ny vil den forbli statisk. Digitalt kan elevene relativt enkelt sette opp ulike representasjoner, og analysere dets egenskaper, samtidig som den initiale representasjonen kan endres raskt etter behov (Sacristán, et al., 2010).

Det hevdes til og med at «*Det er spesielt i forhold til innlæring av begreper disse fordelene kommer klart frem.*» (Bueie, 2011, s. 13). Ved å frigjøre kognitiv kapasitet ved at datamaskinen tar seg av beregninger og visualisering, vil prosedyrekunnskapene ved å eksempelvis konstruere grafer ikke lenger bare handle om manuelle kalkuleringer, men hvordan tolke resultatene og bruke programmet (Sacristán, et al., 2010). Drijvers (2012) viser til gode resultater der bruk av digitale verktøy til opplæring av kalkulus ble brukt til å lære fundamentale konsepter først. I slutten av kurset studentene gjennomførte, ble de undervist i manuelle beregninger. Forskerne kunne peke på at eksamensresultatene til den eksperimentelle gruppen utkonkurrerte en kontrollgruppe som gjennomførte kurset på standardisert måte. Studentene i den eksperimentelle gruppen rapporterte angivelig at «*datamaskinens evne til å ta seg av kalkuleringer og visualisering gav dem selvtillit i arbeidet, og gjorde det mulig å fokusere på den globale problemløsningsprosessen*» (Drijvers, 2012, s. 489).

Kaput (1992) erkjenner videre at i et interaktivt media, som for eksempel Geogebra, vil programvaren kunne reagere på brukerens handlinger uten å være avhengig av en ekstern respons fra lærer eller medelever. I statiske, ikke-interaktive media er derimot brukeren bundet av å selv respondere på det hun har produsert, eller ved de skisserte eksterne aktørene. Digitale teknologier har altså en stor mulighet til å gi tilbakemeldinger på elevarbeid. I tillegg er responsen en får av datamaskinen er det Sacristán et al. (2010) og Hattie (2009) definerer som en upersonlig tilbakemelding. Maskinen gir tilbakemelding uavhengig av hvem eleven er, slik at tilbakemeldinger kan dermed fremstå som mindre truende for elevene som mottar dem (Hattie, 2009). Hvordan en videre kan instruere en datamaskin til å gi tilbakemeldinger av formativ karakter vil derfor adresseres gjennom det påfølgende kapittelet.

2.4 Tilbakemeldinger

Tilbakemeldinger er i mange læringsteorier ansett som et viktig virkemiddel for å influere læring i mange instruksjonskontekster – inkludert digitale læringsmidler (Hattie & Timperley, 2007; Narciss, 2004). I en undervisningskontekst definerer Hattie & Timperley (2007) og Narciss (2013) tilbakemeldinger som all informasjon gitt av en aktør (eks. lærer, medelev, bok, foresatt, eleven selv, erfaring, datamaskin) til en elev om hennes nåværende prestasjon, læring eller forståelse. En lærer kan gi korrigerende informasjon om hvordan en elev har gjennomført en oppgave, en medelev kan fortelle om alternative løsningsstrategier, en bok kan gi ny innsikt i problemet, mens en forelder kan gi oppmuntring. Tilbakemeldinger er altså i følge Hattie & Timperley (2007) en «konsekvens» av prestasjon. For denne oppgavens del vil tilbakemelding som begrep benyttes hovedsakelig i dets kybernetiske format, altså tilbakemeldinger gitt av en maskin.

2.4.1 Formativ tilbakemelding

Effektiv formativ tilbakemelding i matematisk sammenheng har i følge Kulhavy & Stock (1989) to typiske særtrekk – *verifisering* og *utdypning*. Den første av de to kjennetegnene, verifisering, er bekreftelse på hvorvidt et svar er korrekt eller galt. Den vanligste varianten av verifisering innebærer gjerne å markere et svar med «korrekt» eller «galt», et kryss eller en hake, eller annen markering som angir en bekreftelse på hvorvidt svaret faktisk er riktig eller ei. Mer informative verifiseringsmetoder finnes også – deriblant å markere *hvor* i svaret feilen ligger (Shute, 2008).

Graden av utdypning kan ofte variere mer enn de ulike typene verifisering. Valerie J. Shute beskriver i sin litteraturgjennomgang hvordan også denne typen tilbakemelding gjerne kan deles inn i subtyper – en mer *spesifikk* type, eller en *generell* (2008).

2.4.2 Spesifikke utdypende tilbakemeldinger

Spesifikke utdypende tilbakemeldinger kan adressere relevante tema, det korrekte svaret eller forklare hvorfor det gitte elevsvaret er feil. Her responderer tilbakemeldingen eleven får på hva hun selv har gjort i sitt svar. Eksempelvis kan en elev arbeide med å oversette et funksjonsuttrykk til grafisk representasjon, og avgi svaret med tolkningsfeil som resulterer i at uttrykkets konstantledd definerer grafens stigning. En spesifikk utdypende tilbakemelding kan i dette tilfellet forklare at grafens stigning ikke stemmer overens med uttrykket. Den kan videre vise at kilden til dette kan være en misforståelse om hvilket element i uttrykket som korresponderer med grafens stigning.

Shute (2008) hevder at tilbakemeldinger som spesifikt adresserer svaret til elevene ser ut til å øke elevprestasjoner mer enn andre typer feedback. Høyere spesifisitet om hva feilen er, og utdypning av handlingsrommet videre skisseres som en tilbakemeldingsmetode som assosieres med bedredre prestasjoner. Samtidig advares det mot å være for spesifikk i tilbakemeldingsteksten. For lange og kompliserte tilbakemeldinger kan resultere i at eleven overser dem slik at de forblir uleste og dermed ubrukte (Narciss & Huth, 2002; Shute, 2008).

Spesifikke utdypende tilbakemeldinger finner en sjelden blant private aktørers digitale læringssystemer (Murphy et al., 2014). Denne typen tilbakemelding må i et slikt tilfelle ta hensyn til spesifikke aspekter ved elevsvaret som er galt. Det gjør at programmeringen av tilbakemeldingssystem som inkluderer spesifikke utdypende tilbakemeldinger fort blir kompleks, og dermed vanskelig å implementere for et større sett med oppgaver.

2.4.3 Generelle utdypende tilbakemeldinger

Om en ønsker å gi generelle tilbakemeldinger vil utarbeidede eksempler på løsninger til lignende oppgaver eller enkel rettledning mot korrekt svar i form av hinting være tilfeller av dette (Shute, 2008). En generell utdypning kan også indikere hva den riktige responsen bør inneholde uten å ta opp feilen i selve svaret eleven har avgitt. Som eksempel tar jeg utgangspunkt i samme feilsituasjon som jeg beskriver i kapittel 2.4.2. En generell utdypende tilbakemelding kan, i tilfellet av å misforstå en grafs stigning som det tilhørende uttrykkets konstantledd, gi et eksempel på hvordan en lignende funksjon med andre verdier skal oversettes til graf.

Generelle utdypende tilbakemeldinger er blant de vanligste brukte metodene når datamaskinen selv skal bestemme og gi tilbakemeldinger. Ettersom programvaren i et slikt tilfelle ikke trenger å ta hensyn til andre faktorer enn hvorvidt elevsvaret er korrekt eller ei, er det enklere å programmere inn en generell tilbakemelding (Rønning, 2017).

2.4.4 Tilbakemeldinger har ikke nødvendigvis positiv effekt

Det finnes noen farer ved tilbakemeldinger utover at de kan være for lange eller komplekse. Kluger & DeNisi (1996) hevder i sin litteraturgjennomgang at 38% av studier på tilbakemeldinger viser til negativ effektstørrelse på læring. Dette eksemplifiseres med studier på verifiserende tilbakemeldinger innbakt i digitale læringssystemer fra 80- og 90- tallet. Disse studiene sammenlignet utfallet av å benytte seg av verifiserende tilbakemeldinger som markerte hvor i svaret feilen ligger, samt hvorvidt svaret var riktig eller ei, med det å arbeide med det samme læringssystemet uten tilbakemeldinger. Sammenlikningen viser til at disse rent verifiserende tilbakemeldingene hadde en negativ effekt på antall korrekte svar.

Det finnes altså ingen garanti for at automatiserte tilbakemeldinger vil ha en positiv effekt, men det bør nevnes at det gitte eksempelet ikke presenterte utdypende tilbakemeldinger i sin programvare. Artikkelforfatterne påstår dog at tilbakemeldinger på oppgaver i de fleste tilfeller ser ut til å ha en positiv effekt på læring, spesielt om oppgavetypene er kjente for elevene og tilbakemeldingene adresserer elevsvaret med hint mot den konseptuelle kunnskapen og prosedurale ferdigheten som kreves for å mestre oppgaven (Kluger & DeNisi, 1996).

2.4.5 Typer utdypende tilbakemelding i digitale læringssystemer

Utdypende tilbakemeldinger kommer i et stort antall varianter og former i forskningen på tilbakemeldingsfeltet. I enkelte studier fremkommer denne typen tilbakemelding som alle typer tilbakemelding som inneholder mer informasjon enn simpel verifisering (Kulhavy & Stock, 1989; Narciss & Huth, 2002, Shute 2008). Narciss definerer likevel i sin oppsummeringsartikkel fra 2013 hvilke typer utdypende tilbakemeldingskomponenter hun har benyttet seg av i tidligere studier om interaktive digitale læringssystemer. Hun bruker begrepet *informative tutoring feedback-model* for å kategorisere disse komponentene som:

- Kunnskap om oppgavens rammer – gi informasjon om oppgavens «regler» og tydeliggjøring av oppgavens mål
- Kunnskap om konsepter – adressering av den konseptuelle kunnskapen som kreves ved å f.eks. gi repsonsavhengige hint om de relevante konseptenes egenskaper
- Kunnskap om feil – tilby informasjon om feil eller misforståelser (f.eks. spesifisere hvor feilen(e) ligger, eller gi hint om type feil i svaret eller dets kilde)
- Kunnskap om hvordan prosessere oppgaven – adressering av den prosedurale ferdigheten som kreves (f.eks. responsavhengige hint om relevante prosedurale ferdigheter eller problemløsningsstrategier)
- Kunnskap om metakognisjon – tilby informasjon om metakognitiv kunnskap og strategier for selvregulering av læring (f.eks. tema- eller repsonsavhengige hint, se kapitel 2.5)

Disse typene utdypende tilbakemeldingstypene er ikke overaskende sterkt knyttet til de aktuelle oppgavenes tema og relevante nøkkelferdigheter som må forstås for at en elev skal kunne løse dem. Dersom vi skal tro Shute (2008) er utdypende tilbakemeldingers spesifisitet i relasjon til elevresponsen kritisk for å promotere læring. Det neste naturlige spørsmålet blir dermed hvordan en velger ut innholdet i disse tilbakemeldingene.

2.4.6 Automatisert formativ tilbakemelding

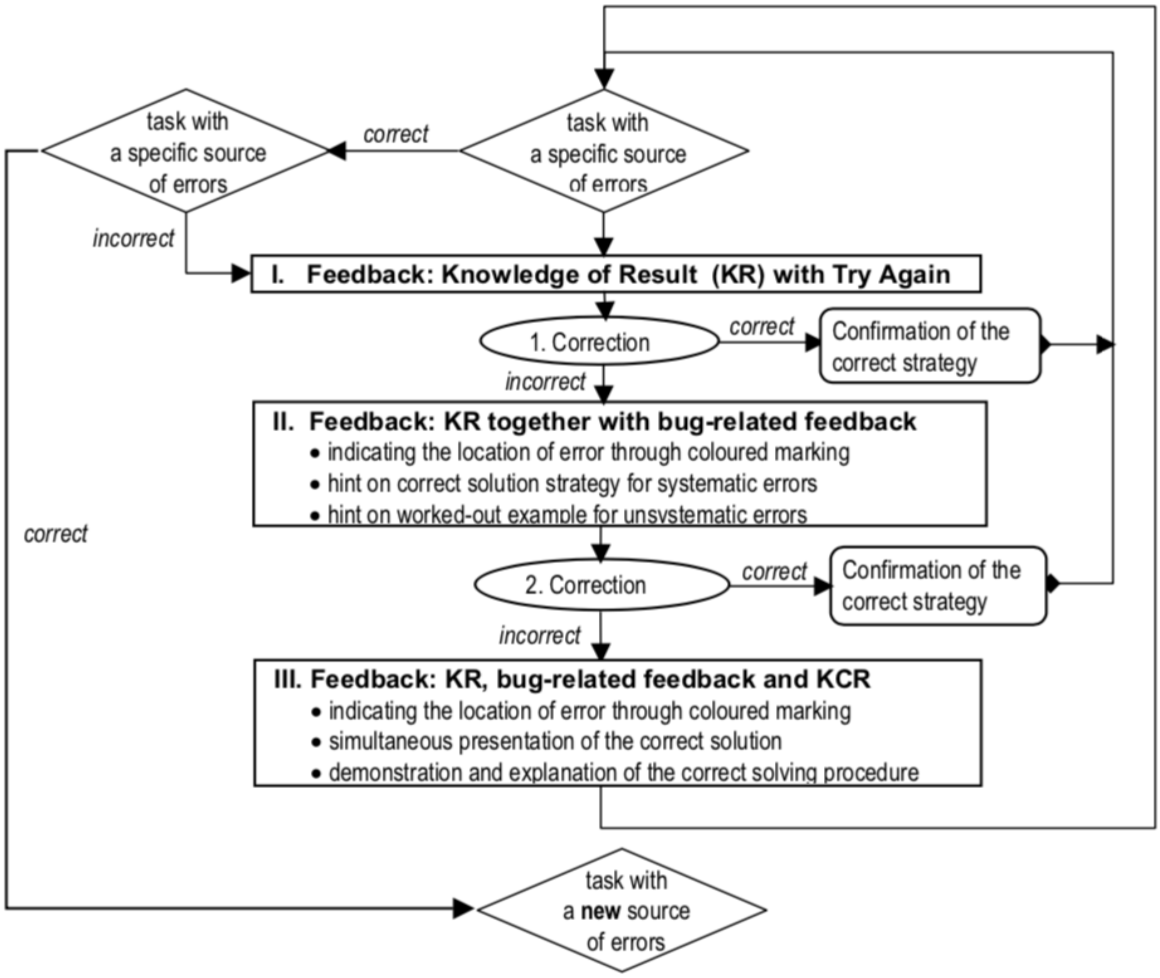
Interaktive digitale læringssystemer har et stort potensial til å gi både verifiserende og utdypende tilbakemeldinger. Et system som kan automatisk vurdere hva eleven har gjort i arbeidet sitt kan utarbeides på mange ulike måter. Med en *algoritmisk* fremgangsmetode for automatisering av tilbakemeldinger, vil en designe ulike oppgaver til elevene, hvorpå forventede elevsvar på forhånd må avklares til hver enkelt oppgave. Dermed kan en programvare kan respondere på elevsvarene kontekstuelt. Ulike typer elevsvar skal altså trigge ulike typer respons fra maskinen, noe som betyr at et stort antall feiltilstander må pre-programmeres inn i programvaren som oppgavene ligger på.

Narciss & Huth (2002) beskriver en slik fremgangsmåte i et forsøk på å designe et automatisk, utdypende tilbakemeldingsverktøy for å lære elever å arbeide med den standardalgoritmen for subtraksjon. Målet med tilbakemeldingene programvaren deres gir er å minske diskrepansen mellom elevens nåværende og ønskede forståelse. For å få til dette mener Narciss & Huth (2002) at tilbakemeldingene må adressere elevsvarets korrekthetsgrad, men også inneholde aspektene slik de fremkommer i *the informative tutoring feedback-model* (Narciss, 2013).

For at tilbakemeldingene på best mulig vis skal kunne gi utdypende informasjon for temaet oppgavene dreier seg om, må også lærings- og kunnskapsmålene for temaet defineres tydelig. Med andre ord hvilke konseptuelle og prosedurale ferdigheter som kreves for å løse oppgavene i programvaren. For å få til dette gjennomførte de en kognitiv oppgave- og feilanalyse for å identifisere de ulike kravene som kan føre til systematiske feil i subtraksjonsoppgaver (Narciss & Huth, 2002; Narciss, 2013). De følger fem steg i en konkretisert prosess der de:

1. Velger og spesifiserer læringsmålene og det ønskede læringsutfallet for det gitte matematiske domenet
2. Velger typiske oppgaver som passer med læringsmålene og læringsutfallene
3. Gjennomfører av kognitiv oppgaveanalyse for å identifisere hvilke domenespesifikke konseptuelle kunnskaper, kognitive operasjoner, prosedurale eller metakognitive ferdigheter som kreves for å mestre oppgavene (kjennskap til subtraksjonsreglene, konseptet null, kjennskap til «låning» av tiere etc.)
4. Gjennomføring av feilanalyse for oppgavesettet for å identifisere typiske systematiske feil, ukorrekte løsningsstrategier, misforståelser eller kilder til typiske feil
5. Velger og spesifiserer informasjonen som skal gis i tilbakemeldingene når elevene møter på feil (hint om feilen eleven har gjort, hint til en metakognitiv strategi, hint til relevant konseptuell kunnskap eller prosedural ferdighet, metakognitivt motspørsmål relativt til elevresponsen, forklaring på en begått systematisk feil etc.)

Resultatet av oppgave- og feilanalysen er deres programvare som implementerer en tredelt flerforsøksstrategi med *bug-related tutoring* (feilrelatert veiledning), og demonstreres gjennom flytdiagrammet i figur 2 på neste side. Programvaren med feilrelatert veiledning kan beskrives ved at elev har tre muligheter til å klare hver enkelt oppgave. Typen respons fra programmet endrer seg avhengig av hvor mange forsøk eleven har brukt på oppgaven og hvorvidt samme type feil gjentar seg eller ei. Jo flere feil eleven har brukt, jo mer spesifikk blir utdypningen. Etter første forsøk får eleven presentert hvorvidt det avgitte svaret er korrekt eller ei sammen med en beskjed om å prøve på nytt. Ved andre forsøk gis verifisering av svaret sammen med farget markering av hvor feilen er, og med enten et hint mot løsningsstrategi dersom programvaren oppdager at feilen er av samme type som ved første forsøk, eller et eksempel på løsning av liknende oppgave. Ved tredje forsøk gis verifisering av svaret sammen med farget markering av hvor feilen er, presentering av det korrekte svaret samt løsningsstrategi og forklaring av denne.



Figur 2, flytdiagram for feilrelatert veiledning (Narciss & Huth, 2002, s. 12)

Narciss & Huth (2002) og Narciss (2013) gir også noen tydelige anmodninger om hvordan disse tilbakemeldingene skal presenteres, i hvilken rekkefølge og hva de *ikke* bør inneholde. De hevder at tilbakemeldinger ikke bør gis før eleven selv har forsøkt seg på en oppgave, da spesielt ikke det korrekte svaret på oppgaven. Videre anbefaler de å gi utdypende tilbakemeldinger i stegvis økende presisjon eller kompleksitet, for å unngå at informasjonen i tilbakemeldingen blir oversett eller for vanskelig å forstå. Denne implementeringen krever at elevene har flere forsøk på én og samme oppgave. For programvare som skal brukes som læringsverktøy over lengre tid oppfordres det til å implementere et slags «mestringsnivå» i programvaren. I praksis betyr dette at eleven ikke får nye typer oppgaver før et konsept eller en prosedural ferdighet er mestret.

2.5 Metakognisjon og selvregulert læring i matematikk

Andre aktører forsker på tilbakemeldingers rolle i undervisningskontekster gjennom sosio-kognitive øyne, og har benyttet seg av et større fokus på metakognisjon og selvregulering av atferd. Bracha Kramarski og Zemira Mevarech har forsket i over tyve år på selvreguleringens rolle innenfor læring av matematikkfaget. Gjennom flerfoldige artikler leveres hovedbudskapet om hvordan elever selv har et ansvar for å ta kontroll over egen læring (Kramarski & Zeichner, Using Technology to Enhance Mathematical Reasoning: Effects of Feedback and Self-Regulation Learning, 2001; Mevarech & Kramarski, 2003). Metoden de er kjent for å bruke – enten det innebærer læring gjennom digitale midler eller ei, er en undervisningsmetodikk som blant annet dreier seg om å stille metakognitive spørsmål (MKS) som respons på skriftlige eller verbale elevsvar. Når elevene har arbeidet med en oppgave på egenhånd, diskutert et tema i klassen, svart på en prøve eller lignende anbefaler Kramarski og hennes støttespillere å stille slike kognitive motspørsmål.

De metakognitive spørsmålene kan i følge Mevarech & Fridkin (2006) deles inn i fire kategorier; *forståelsesspørsmål*, *sammenhengsspørsmål*, *strategiske spørsmål*, og *refleksjonsspørsmål*. Generelt vil en lærer som utøver denne metodikken snakke om disse spørsmålene når et introduseres et nytt tema og bruke dem i tilbakemelding til elevene. I tillegg er det ønskelig at elevene snakker rundt disse spørsmålene når de forklarer sin matematiske problemløsing. De fire kategoriene av spørsmål handler konkret om:

1. [...] det å forstå problemet (f.eks. «Hva handler dette problemet egentlig om?»);
2. [...] det å konstruere sammenhenger mellom tidligere ervervet og ny kunnskap (f.eks. «Hva er likhetene/forskjellene mellom problemet for hånden og de du har løst tidligere?»);
3. [...] det å bruke egnede strategier for å løse problemet (f.eks. «Kan du tenke deg en **strategi**, **teknikk** eller **metode** du kan bruke for å sette opp grafen?»; «Hva trenger du?»)
4. [...] refleksjon rundt selve prosessen og løsningen («Hva gjorde jeg galt her?»; «Gir løsningen jeg har funnet mening?»)

(Kramarski & Gutman, 2006, s. 25)

Det foreslås at denne måten å svare elevene med flere spørsmål bør implementeres som en del av undervisningspraksis med syklisk og rekursiv gjentagelse. Hensikten er at elevene skal kunne bli bevisste over sin egen læring og derfor bli kapable til å endre sine læringsstrategier.

I forbindelse med tidligere forskningsprosjekt lot hun denne typen tilbakemeldinger presenteres gjennom en datamaskin og sammenlignet testresultatene opp mot en kontrollgruppe som fikk verifiserende tilbakemeldinger. Konklusjonen er ifølge forskerteamet at elever som blir utsatt for læring orientert rundt metakognitive spørsmål og selvregulert læring presterer bedre i tester enn motparten som får de rent resultatorienterte tilbakemeldingene. Spesielt stor effekt opptrådde når programvaren gav tilbakemelding i form av metakognitive spørsmål som fokuserte på forskjeller og likhet mellom oppgaver. Å forsøksvis forstå et problem før en prøver en løsning, samt refleksjon over hvilke strategier og metoder som kan lede mot en løsning var virkningsfulle (Kramarski & Zeichner, 2002).

2.6 Nettbaserte math.ED- verktøy og tilbakemeldingssystemer

Math.ED- verktøy, eller matematiske e-læringsverktøy er basert på ideen om at elever kan få tilgang på læringsressurser og aktiviteter i deres eget tempo i forhold til deres behov (Barana & Marchisio, 2016). I tillegg til forskning på feltet, slik det eksemplifiseres i kapittel 2.3.2 og 2.3.3, finnes det en rekke private aktører som leverer ulike nettbaserte plattformer og vurderingsverktøy for matematisk læring. Jeg vil i dette delkapittelet utdype hvordan noen av de mer populære nettplattformene fungerer med tanke på deres tilbakemeldingssystemer. Enkelte av dem, som Khan Academy og Gyldendals Maximum Smartøving, trakter ikke bare vurdering av matematikkoppgaver, men har et spesielt fokus på å produsere læringsinnhold i form av eksempler på oppgaver med løsningsmetoder, videoklipp og tekstuelt innhold. De har dog noe forskjellig fokus på ulike aspekter ved sitt innhold, bruker ulike metoder med tanke på hvordan deres oppgaver er satt sammen og på hvilken måte løsninger automatisk vurderes av de respektive systemene.

Forskningen på feltet innen automatiske tilbakemeldingssystemer kan fortelle oss om konsekvensene av bruken av slike verktøy. På den annen side utsettes elever i skolen for i all hovedsak ikke for forskning, men for de private aktørenes produkter. I tillegg setter det min egen undersøkelse i en kontekst av samtidens e-matematikklæringsprodukter. Det blir i så måte nødvendig å undersøke hvordan disse produktene fungerer relativt til deres tilbakemeldingssystemer.

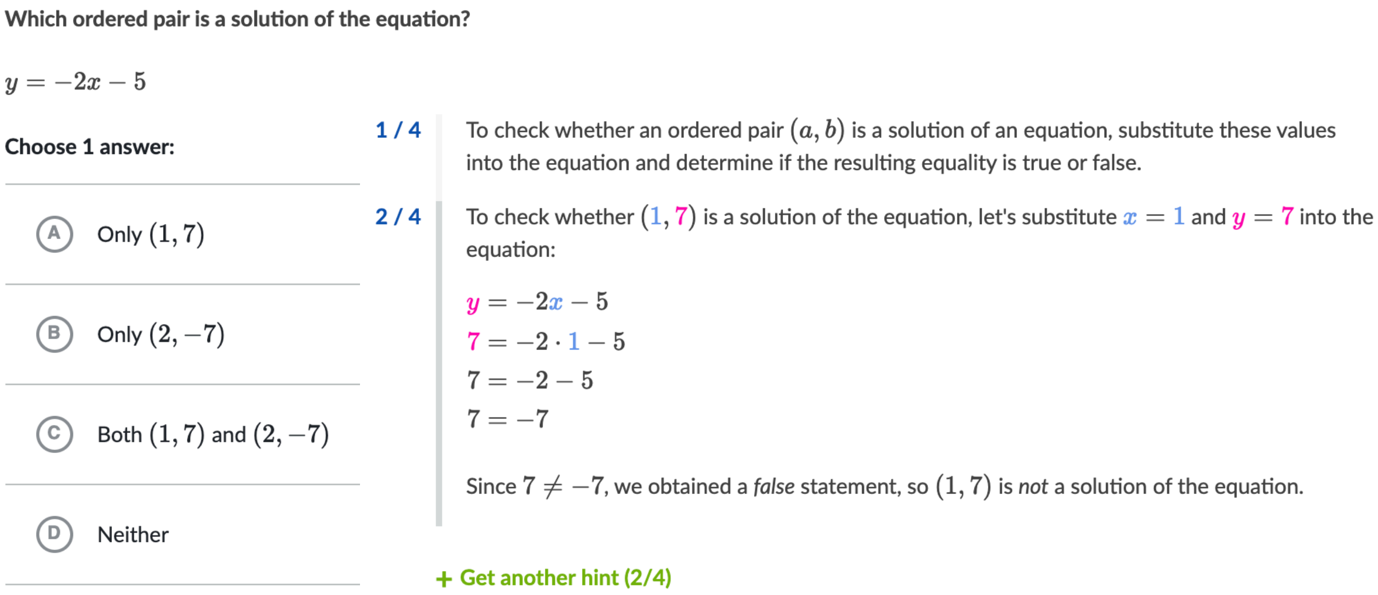
Det finnes en rekke masteroppgaver, deriblant Ødegaard (2016) og Dahl (2014) og en rapport (Egelandsdal et al., 2019), som beskriver henholdsvis Maximum Smartøving og Kikora som produkter. I tillegg har en rapport etter Oslo kommunes evaluering av Kikora på bestilling fra Utdanningsetaten blitt publisert (Flaa, 2009), samt SLATE i Bergen sin FoU-rapport av Morlandstø, Hansen, Wasson, & Bull (2019) etter bestilling av KS blitt vektlagt for disse plattformene. Andre publiserte forskningsartikler om bruken og dets konsekvenser relatert til Maple T.A. presenteres også (Barana & Marchisio, 2016; Char, 2011; Rønning, 2017).

2.6.1 Khan Academy

Khan Academy er sammen med Maple T.A. en de to internasjonale aktørene jeg nevner i dette delkapittelet. De er også den største e-læringsplattformen i global skala, mye takket være det faktum at alt innhold de leverer er gratis, samt at innholdet er oversatt til 42 språk. Plattformen ble først kjent for korte videoklipp som ble lagt ut på YouTube i 2006. Her instruerer deres grunnlegger, matematiker og informatiker Salman Khan, i regnemetoder og matematiske konsepter som undervises i ungdomsskolen. Tjenesten ble etterhvert ekspandert med mengder av øvingsoppgaver i mange matematiske tema fra barneskole til videregående opplæring. Plattformen tilbyr innhold for lærere som gjør det mulig å følge deres elevers progresjon og tidsbruk gjennom ulike tema.

Oppgavene man finner på nettsidene deres er typisk av flervalgsvarianten, men de har også oppgaver der elever kan konstruere grafer i koordinatsystem, samt oppgaver som krever manuell beregning som kan besvares med variabler, likninger og tallmessig format (Murphy, Gallagher, Krumm, Mislevy, & Hafter, 2014). Oppgavene anbefales å gjennomføres i en bestemt rekkefølge sammen med videoinnhold, men elevene kan fritte velge hvor de ønsker å begynne. Man finner også situasjonsbeskrivende oppgaver, men også her svarer eleven gjennom de overnevnte formatene.

Khan Academy håndterer tilbakemeldinger til elevene på tre ulike måter. Den første måten, hvor en elev ved galt svar får en verifiserende feilmelding opp på skjermen, sammen med en beskjed om å prøve på nytt (Murphy et al., 2014). Måte nummer to dukker opp dersom en svarer feil mer enn tre ganger på rad. I dette tilfellet vil nettsiden opplyse om at eleven kan trykke på en «hint»-knapp eller se et videoklipp som omhandler det relevante temaet. Hint-knappen er forøvrig alltid tilgjengelig, og kan brukes så ofte eleven føler for. I praksis fører det til at hintene gir løsningsforslag til oppgaven for eleven, med forklaringer om hvilke prosedyrer som skal benyttes, slik en kan se i figur 3 under. Sammenlignet med Narciss (2013) og Narciss & Huth (2002) sine anbefalinger, plasserer altså Khan Academy sin tilbakemeldingsmetode seg inn i den mest førende varianten av utdypende og verifiserende tilbakemeldinger.



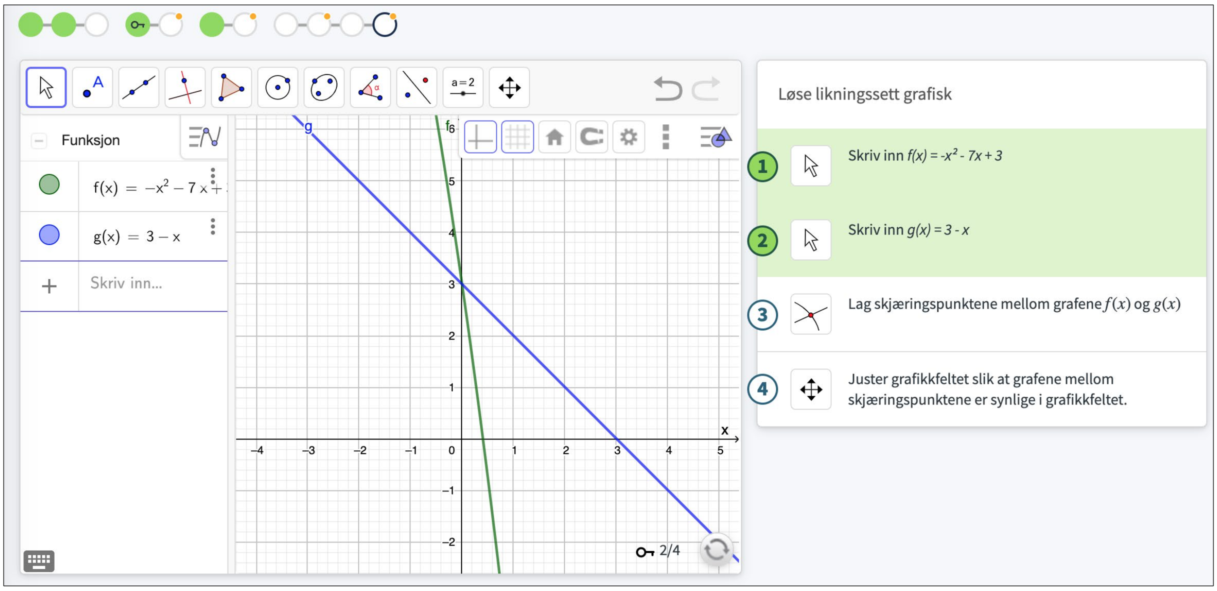
Figur 3, på venstre hånd vises en flervalgoppgave, mens resultatet av å trykke på hint-knappen vises på høyre hånd (fra khanacademy.org)

Den tredje måten de serverer tilbakemeldinger på kommer i form av virtuelle medaljer og «energipoeng» som belønning for gjennomføring av spesielle oppgavesett. Når en viss mengde energipoeng er nådd kan brukeren øke sitt virtuelle nivå på nettsiden. Disse eksterne motivasjonsfaktorene kan ha blitt påvirket av både behavioristisk tankegods og «gamification». Det vil si at læringsverktøyet er influert av spilldesign for å motivere elever til å jobbe med læringsaktiviteter (Landers, 2015).

2.6.2 Kikora

Kikora er en norskutviklet nettplattform som i hovedsak har tilbudt digitale matematikkoppgaver på nett siden 2008. Av spesielle fordeler med å benytte den plattformen finner jeg at Kikora har implementert en mulighet til å arbeide med Geogebra gjennom deres nettside.

Oppgavene spenner seg mellom rene flervalgoppgaver og oppgaver som manuelt må beregnes, til oppgaver basert på å følge instrukser på skjermen. Et sentralt aspekt ved Kikora er at elevene kan fortløpende få tilbakemeldinger på hvorvidt løsningene deres er korrekte eller ei (Flaa, 2009; Morlandstø et al., 2019). Det gir elevene mulighet til å raskt gå tilbake i et oppgavesett og rette opp sine feil. I tillegg kan elevene få en delvis eller hel løsning av hver enkelt oppgave ved å trykke på en nøkkelknapp, slik det fremkommer i figur fire. Dersom en elev bruker løsningsforslaget, gjør Kikora regelrett oppgaven for elevene. Morlandstø et al. (2019), påpeker at ca. 11% av brukerne benytter seg av løsningsforslag, en relativt liten andel.



Figur 4, her vises en oppgave om funksjoner og grafer fra på Kikoras nettside, hvor to av delmålene i oppgaven er løst av programvaren ved å trykke på nøkkel-knappen (med tillatelse fra kikora.no)

Et interessant funn de har gjort handler om antall steg elevene bruker per oppgave. Gjennomsnittlig benytter hver bruker 2,1 steg – noe Morlandstø et al. (2019) hevder kan vise til at oppgavene innad i Kikora er for enkle. De forteller videre at elever ser ut til å få noe bedret læringsutbytte av å benytte seg av Kikora, mye på grunn av at ordlyden og språket som brukes i oppgavene sammenfaller med det en finner i klassiske lærebøker (Morlandstø et al., 2019). Elevene får altså umiddelbar verifiserende tilbakemelding, men ikke utdypende om hva som er galt, slik Narciss & Huth (2002) foreslår i sin artikkel. Den rent formative tilbakemeldingen uteblir dermed i Kikora sin plattform. Repetisjon og mengdetrening ser ut derfor ut til å være bruksområdene der Kikora egner seg best (Dahl, 2014). I tillegg gir Kikora, i likhet med Khan Academy, virtuelle medaljer og trofeer for å løse spesielle oppgavesett. Disse trofeene kan vises frem på brukerens side, og gis som en slags motivator og belønning for å jobbe med Kikoras nettside.

2.6.3 Gyldendal Maximum Smartøving

Gyldendal har valgt seg ut en litt annen tilnærming til e-læringsverktøy enn Kikora og Khan Academy, som tilbyr det jeg vil beskrive som mer statiske plattformer. Den store forskjellen er at programvaren til Gyldendal forsøker å adaptivt tilpasse oppgavesettet til elevene avhengig av deres prestasjoner (Egelandsdal et al., 2019). Slike systemer kalles gjerne for Adaptive Læringsplattformer (AL), eller Intelligent Tutoring Systems (ITS) (Egelanddal et al., 2019; Hattie, 2011). Om deres produkt Maximum Smartøving, som er myntet på ungdomstrinnene sier de følgende: «*Multi Smart Øving bygger på avansert teknologi som gjør at systemet avdekker hva elevene kan og ikke kan. Dersom en elev gjør gjentatte feil vil systemet forsøke å avdekke eventuelle manglende forkunnskaper, og gi eleven oppgaver for å bøte på manglene. Hvis elevene likevel fortsetter å gjøre de samme feilene, vil programmet finne en video eller et eksempel som forklarer læringsmålet.*» (Gyldendal Akademisk, 2019).

En plattform som kontinuerlig tilpasser sine oppgaver basert på elevenes prestasjoner, på denne måten omtales av og til i litteraturen som et «Intelligent Tutoring System» (Hattie, 2011; Narciss, 2013). I likhet med Kikora og Khan Academy er oppgavene som Gyldendal presenterer stort sett av flervalgstypen, eller oppgaver som krever en manuell beregning hvorpå svaret tastes inn i et svarfelt. Tilbakemeldingene i Maximum Smartøving fremstår her som mer utdypende og verifiserende. Ettersom nettsiden kontinuerlig prøver å tilpasse seg det den tror er elevens nåværende nivå for å deretter gi nye oppgaver tilpasset dette nivået, kan det dermed sies at programvaren til en viss grad forsøker å gjøre en del av vurderingsjobben en lærer typisk vil ha. Da Maximum Smartøving også tilbyr videoklipp og eksempler etter det antatte nivået til eleven forsterkes denne påstanden ytterligere.

I tillegg får læreren kontinuerlig oversikt over hver elevs progresjon og arbeidstid på hvert enkelt tema. Ødegaard (2016) påpeker at systemet ser ut til å øke elevenes tidsbruk på det matematiske arbeidet gjennom bruk av programvaren. Samtidig advares det mot at en slik løsning kan videre stimulere elevene til et prestasjonsorientert fokus. Plattformen gir elevene en poengsum eller skår som kan medføre et større prestasjons- og konkurranseorientert fokus fremfor et mer anvendelig konseptorientert et (Egelandsdal et al., 2019; Ødegaard, 2016).

2.6.4 Maple T.A.

Maple T.A. er et vurderingssystem som skiller seg ut fra de øvrige plattformene jeg beskriver over i sin kraft av å være et automatisk summativt vurderingssystem først og fremst (Barana & Marchisio, 2016; Rønning, 2017). Generelt vil summativ vurdering ha som formål å kartlegge elevens eller studentens nåværende kompetanse. Denne typen vurdering benyttes altså for å godkjenne eller sertifisere enn person til et videre formål. Flere universiteter i Europa og USA benytter seg av Maple T.A. for å godkjenne øvingsopplegg og arbeidskrav. Systemet gjør det mulig for undervisere å lage, administrere og auto-vurdere nettbaserte oppgaver (Char, 2011; Rønning, 2017). Maple T.A. har i likhet med Kikora en verifiserende og dikotomisk vurderingsmetodikk. Korrekte svar markeres med en grønn farge i en tydelig melding om at eleven har løst oppgaven. På samme vis markeres gale svar med rød farge og en verifiserende feilmelding, altså «galt svar». Rønning (2017) priser Maple T.A. for evnen det har til å raskt kunne bedømme om studentene i et fag kvalifiserer seg til å gå opp til eksamen, men retter likevel advarsler mot å belage seg for sterkt på programvaren. Ved å verifisere svar sammen med en «prøv-til-du-er-ferdig» metodikk hevder Rønning (2017) at det er en overhengende fare for at studentene vil gjette seg fram i blinde etter mange nok forsøk, i en prosess han kaller «jakten på svaret». «*The fact that only the final answer is evaluated, and only with right or wrong, seems to encourage a way of working that, at least after some failures, pays little attention to how the answer is obtained and turns into a "random hunt for the correct answer"*» (Rønning, 2017, s. 104).

Som det er mulig å skimte gjennom beskrivelsene i dette delkapittelet plasserer de private aktørenes tilbakemeldingssystemer seg i ulike kategorier. Likevel kan det argumenteres for at flere av dem, spesielt Khan Academy og Kikora sine løsninger, muligens er noe raske i å gi verifiserende tilbakemeldinger i form av løsninger og løsningsforslag. Oppgavene er også av den mer prosedyreorienterte varianten, slik at modellering og problemløsningsoppgaver fort kan utebli.

3 Metode

Undersøkelsen søker å belyse hvordan automatisk maskingitte tilbakemeldinger på digitale matematikkoppgaver i representasjoner av funksjoner kan utformes, og hvordan tilbakemeldingene påvirker elevene som interagerer med systemet. Jeg vil derfor minne om forskningsspørsmålene til undersøkelsen:

1. Hvordan kan det designes automatisk datamaskingitte, utdypende og verifiserende tilbakemeldinger til oppgaver om oversettelse fra tabellform, algebraisk- og situasjonsrepresentasjon til grafisk representasjon av lineære funksjoner?
2. Hvordan påvirker automatiske verifiserende og utdypende tilbakemeldinger gitt av en datamaskin elevers prestasjoner i oversettelsesoppgaver i lineære funksjoner?

Det første delspørsmålet vil besvares gjennom metodekapittelet, mens resultatene av de ulike tilbakemeldingenes påvirkning fremkommer i analysekapittelet. I det påfølgende kapitelet vil jeg beskrive metodologiske og metodiske valg som er gjort med tanke på forskningsdesign, datainnsamling, utforming av testen og dens oppgaver, tilbakemeldinger, samt analyseprosessen. Innledningsvis vil jeg skissere hvilket forarbeid som er lagt til grunn med tanke på gjennomføring av selve datainnsamlingen i undersøkelsen. Dernest vil jeg beskrive elevutvalget som er inkludert før min metodologi *semi-randomisert eksperiment* vil presenteres. Videre utdyper jeg hvordan den kvantitative datainnsamlingsmetoden i form av en *test* har vært gjennomført, samt hvordan oppgavene i testen er utformet og hvilke spesielle utfordringer elever kan møte i disse. Utformingen av tilbakemeldingene elvene får i arbeid med oppgavene vil deretter diskuteres før erfaringer fra gjennomføringer av pilotundersøkelsen og den reelle datainnsamlingen drøftes. Til slutt vil analysemetoden samt forskningens troverdighet og et delkapittel om metodekritikk tas opp.

3.1 Forarbeid til undersøkelsen

For å avgjøre om det var mulig å gjennomføre undersøkelsen ble flere matematikklærere på 10. trinn ved ungdomsskoler i Trondheimsregionen kontaktet tidlig i arbeidsprosessen. Hos lærere som fattet interesse ble en potensiell gjennomføring av undersøkelsen avklart med ledelsen ved skolen. Matematikklærerne ble i forkant bedt om å svare på hvor godt kjent deres elever er med Geogebra eller lignende programvare og hvordan de bruker det i undervisningsøyemed. I tillegg ble de forespurt om deres elever hadde arbeidet med tema funksjoner i løpet av semesteret datainnsamlingen foregikk, og når de eventuelt skulle gjøre det i fremtiden.

For å gjennomføre datainnsamlingen ble det digitale testverktøyet Matistikk brukt. Nettplattformen er under utvikling av Institutt for lærerutdanning ved NTNU og benyttes til å samle inn matematiske besvarelser på internett. En stor fordel ved å benytte seg av en slik datainnsamlingsplattform er anonymiteten det gir elevene som svarer på oppgaver gjennom det. Ingen personidentifiserende data blir innsamlet og testen kan enkelt administreres gjennom de vanligste nettleserne Google Chrome, Apple Safari, Opera, Mozilla Firefox og Microsoft Edge.

I tillegg kan Matistikk programmeres slik at det sporer aktiviteten til testdeltagerne i arbeid med oppgavene. Sporingen fremstår som et kritisk element i denne sammenhengen ettersom det legger til rette for at visse tilstander og egenskaper i grafikkfeltet kan evalueres av en algoritme.

3.1.1 Elevutvalget

Elevutvalget består av totalt 123 10. klasseelever fordelt på fem klasser over fire skoler i Trondheim. Disse 10. klassene ble valgt ut på grunnlaget av å ha hatt sin siste undervisning i funksjoner på omtrent samme tid. Alle de fem klassene hadde sist jobbet med funksjoner på skolen ved slutten av foregående semester, altså i mai og juni. Dette betyr at det var fem til seks måneder siden de sist hadde arbeidet med temaet de ble testet i: nemlig lineære funksjoner.

3.2 Beskrivelse av undersøkelsen

I denne undersøkelsen blir de deltagende elevene utsatt for et eksperiment der de gjennomfører en matematisk test. Ved å sette opp et eksperiment får jeg muligheten til å teste hvordan en uavhengig variabel, utdypende tilbakemeldinger på oppgavene i testen, påvirker en avhengig variabel, altså prestasjonene i testen på tvers av gruppene. Dermed må eksperimentering som metodologi, mer spesifikt randomiserte forsøk defineres.

Eksperimenter har en lang historie i naturvitenskapen. Undersøkelser av naturfenomener låner seg gjerne godt til eksperimentering, ettersom det rasjonalistiske verdisynet det hviler på tillater forskeren å fragmentere virkeligheten inn i mindre biter som videre kan kvantifiseres, analyseres og undersøkes (Guba, 1981).

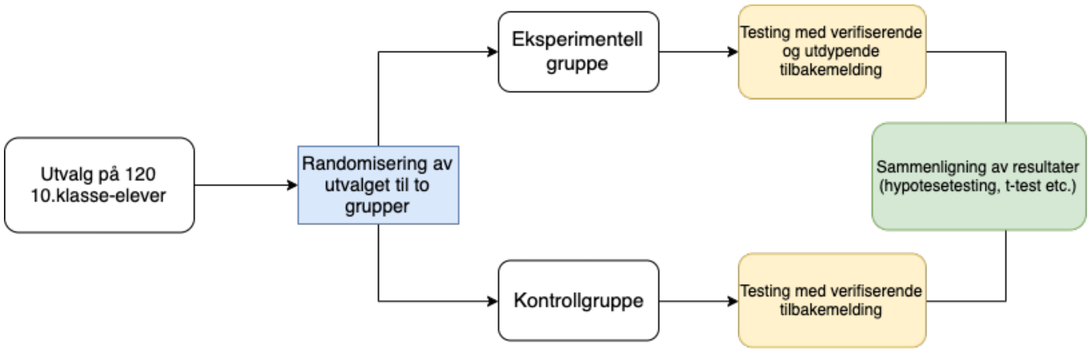
3.2.1 Eksperimentering - randomisert kontrollert forsøk som kvantitativ metodologi

Dersom en vil adressere kausalitet og effekt av konkrete endringer på et utvalg vil eksperimenter kunne bidra i letingen på svar (Cohen, Manion & Morrison, 2018). Når begrepet *eksperiment* dukker opp i forskningslitteratur, finner man gjerne at forskerne snakker om et *randomisert kontrollert forsøk,* (RKF) (Cohen et al., 2018; Pontoppidan et al., 2018; Styles & Torgerson, 2018,). Et RKF søker å avsløre sammenhengen mellom en intervensjon og dets effekt – altså identifiserer den *hva* som virker og ikke virker, men den vil ikke nødvendigvis kunne fortelle forskeren *hvorfor* (Cohen et al., 2018). Når en gjennomfører et eksperiment, ønsker forskeren å endre én enkelt variabel, den uavhengige variabelen, for å så observere dets effekt på en annen, den avhengige variabelen. Trikset er å sammenligne utfallet av en gruppe som blir utsatt for en intervensjon med en kontrollgruppe som ikke får denne uavhengige variabelen påvirket.

Den kritiske faktoren i et eksperimentelt forsøk er forskerens evne til å kontrollere ulike forhold som kan påvirke et resultat. Et slikt forhold for RKF er randomisering, både når det gjelder utvelgelse av deltagere fra en populasjon, men også når det gjelder hvilken gruppe deltagerne plasseres i, for å motvirke seleksjonsbias (Cohen et al., 2018; Hutchinson & Styles, 2010. Ideelt sett vil et tilfeldig utvalg hentes fra en populasjon. Utvalget vil så tilfeldig sorteres i to grupper, eksperimentgruppen og kontrollgruppen, som videre pre-testes for å sikre paritet mellom gruppene, altså at det ikke er statistisk signifikante forskjeller mellom dem (Cohen et al., 2018; Hutchinson & Styles, 2010). Ved å sette et sammenligningsgrunnlag før intervensjonen utføres, vil en genuin forskjell mellom gruppene bli mindre tilslørt. Etter pre-testen, vil eksperimentet settes i gang – intervensjonen utøves på eksperimentgruppen enten før eller under post-testen. Til slutt sammenlignes resultatene, slik at korrelasjon eller kausalitet kan bestemmes.

For masterprosjektet kan figur 6 på neste side demonstrere hvordan den simpleste varianten av RKF ble gjennomført, og er tatt etter inspirasjon av Hutchinson & Styles (2010). Den har flere av kjennetegnene til et ekte eksperiment, men den mangler bruk av pre- og posttest for å sikre at forsøksdeltagerne i utgangspunktet er sammenlignbare.

123 elever fordelt på fem klasser deltok, der hver klasse ble tilfeldig inndelt i en eksperimentell gruppe og kontrollgruppe. Det er altså viktig å presisere at *hver enkelt klasse* ble delt i to. Alle klassene har dermed representanter for både eksperimentgruppen og kontrollgruppen, noe som gjør det lettere å forsvare en sammenligning av de to gruppene i ettertid. Etter gruppeinndeling, tok de to gruppene den samme matematiske testen om oversettelse av representasjoner av lineære funksjoner. Den eksperimentelle gruppen fikk verifiserende og utdypende tilbakemeldinger underveis i testen, mens den eksperimentelle gruppen kun fikk verifiserende tilbakemeldinger.



Figur 6 - modell for mitt design-eksperiment (egen produksjon)

Effekten av intervensjonen, de utdypende tilbakemeldingene, vil kunne spores gjennom å sammenligne elevenes testresultater på hver enkelt oppgave og i testen som helhet. Dersom en gruppe presterer bedre enn den andre, vil prestasjonsforskjellen kunne kvantifiseres (Cohen et al., 2018; Hutchison & Styles, 2010). På denne måten kan forskningsspørsmålet som dreier seg om hvordan elevenes prestasjoner blir påvirket av å bli utsatt for automatiske utdypende tilbakemeldinger besvares.

Uten pretest bør i det minste noen kriterier for deltagerne bestemmes. I mitt tilfelle vil det å sikre seg at tidspunktet for siste undervisning i lineære funksjoner og paritet i aldersgruppen. Populasjonen jeg henter utvalget mitt fra er altså 10. klassinger som alle hadde siste undervisning av lineære funksjoner på cirka samme tid.

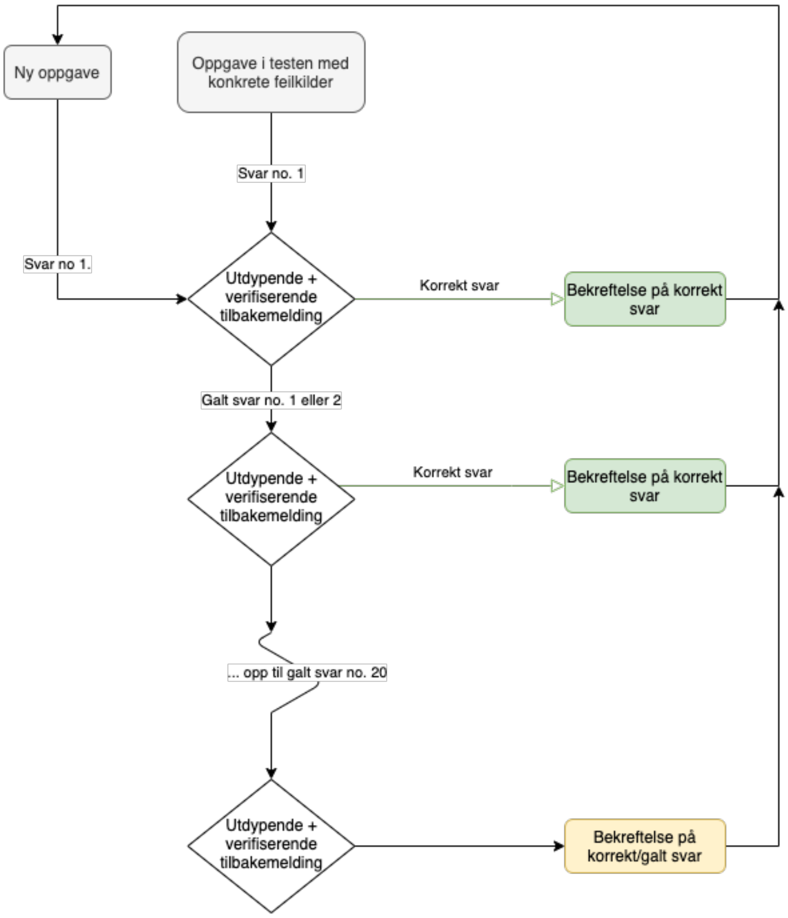
Grunnen for utelatelse av en pre-test er av tidsbegrensning i et masterprosjekt og valgmuligheter. Ved å gjennomføre en pre-test må samtlige testdeltagere testes to ganger. Når det i utgangspunktet er vanskelig nok å rekruttere forsøksvillige klasser til bare én enkelt test for et masterprosjekt, setter det begrensninger for fullstendig randomisering. For undertegnedes masteroppgave kan det stilles spørsmål ved hvorvidt randomisering i utvelgelse av forsøksdeltagere i det hele tatt er mulig. I håp om å randomisere denne prosessen mest mulig ble samtlige ungdomsskoler i Trondheim tilsendt en mail med tilbud om deltagelse. Hvilke skoler som da velger å delta med sine elever som forsøkskaniner er nok ikke helt tilfeldig. I min erfaring var skoler og lærere som tidligere hadde deltatt i forskningsprosjekter mer tilbøyelige til å svare på den utsendte mailen enn andre skoler. Ettersom deltagelse er frivillig kan det bety at lærere og skoler hvor ledelsen er spesielt opptatt av forskning og eksperimentelle læringsmetoder i større grad vil delta. I så måte vil et eksperiment som henter ut deltagere på dette viset, men ikke randomiserer gruppene være et *kvasi-eksperiment* (Cohen et al., 2018). Ordet «kontrollert» i RKF vil også gjerne referere til et pre-posttest-design, noe som gjør at den skisserte metodologien i denne undersøkelsen vil heller kunne kalles et *semi-randomisert eksperiment* (Robson, 2002).

3.2.2 Testing som metodisk datainnsamlingsverktøy

Jeg forsøker å få tilgang til elevenes funksjonskompetanse gjennom en test i oversettelse av funksjonsrepresentasjoner til grafisk representasjon. Tester beskrives som en svært vanlig datainnsamlingsmetode for å få tilgang på informasjon, og det finnes en rekke forskjellige typer (Cohen et al., 2018; Creswell, 2018). I praksis kan testen jeg har konstruert beskrives som en prestasjonsundersøkende test – det vil si at den gjennom oppgavene den inneholder undersøker hva en elev kan eller ikke kan gjøre. Den inneholder fem oppgaver, der riktig svar skåres som 1, og galt som 0, altså er den også en test som samler interval-data (Cohen et al., 2018).

Oppgavene ble designet i det dynamiske geometriprogrammet Geogebra. Filene for hver enkelt oppgave ble modifisert med et enkelt program slik at de kunne sende ut elevenes svarstatus til Matistikk. Svarstatusen eleven har i arbeid med oppgavene leder til at programmet velger ut en bestemt tilbakemelding som skal gis.

Eksperimentgruppen får, som nevnt tidligere, en test med verifiserende og utdypende tilbakemelding, mens kontrollgruppen får en test med kun verifiserende tilbakemelding på deres besvarelser. I motsetning til Narciss & Huth (2002) som har valgt tre-trinns svarmodell, har jeg i denne testen valgt å la elevene kunne svare på hver enkelt oppgave 20 ganger. Etter tjuende forsøk «låser» oppgaven seg, og eleven blir så bedt om å gå videre til neste oppgave. Den hovedsakelige strukturen for tilbakemeldingene til den eksperimentelle gruppen kan sees i figur 7.



Figur 7, enkelt flytdiagram for tilbakemeldingsstrukturen med oppgavene (egen produksjon).

For å beskrive gangen i testens oppbygging for eksperimentgruppen, kan en se for seg at en elev arbeider med en oppgave i testen. Eleven trykker på «check answer»- knappen i testen og finner ut at svaret hun har produsert er galt. Hun får deretter utdypende og verifiserende tilbakemelding og har dermed brukt opp ett av tjue svarmuligheter. Ved neste svarforsøk får hun korrekt svar, hvorpå en melding bekrefter at svaret er korrekt, før oppgaven låser seg. Først nå kan eleven gå videre til neste oppgave. Eleven kan ved et hvilket som helst tidspunkt bestemme seg for å gå videre til neste oppgave uavhengig av å ha klart oppgaven.

En annen vesensforskjell mellom denne undersøkelsen og Narciss & Huth (2002) sin metode er at programvaren aldri gir elevene det korrekte svaret eller løsningsstrategi på oppgaven etter tredje forsøk på svar. Ettersom testen de jobber innenfor ikke skal brukes som et aktivt læringsverktøy over tid, var det en fare for at visning av korrekt svar ville være for førende for elevene og dermed farge resultatene i forsøket.

3.2.3 Læringsmål for oppgavene

Narciss & Huth (2002) og Narciss (2004; 2013) anbefaler forskere som ønsker å utvikle digitale systemer som gir tilbakemeldinger basert på feilrelatert veiledning, å utføre en kognitiv oppgave- og feilanalyse. Oppgave- og feilanalysen hjelper en å bestemme innholdet i og kriteriene for å trigge ulike typer tilbakemeldinger, som igjen er basert på oppgavene i testen. En slik analyse skal inneholde en definering av læringsmål og konkrete læringsutfall, valg av oppgaver som passer med læringsmålene og læringsutfallene, definering av nødvendige konseptuelle kunnskaper og prosedurale ferdigheter, identifisering av typiske feil elever kan gjøre og kilden til disse og til slutt valg av informasjon i tilbakemeldingene (Narciss & Huth, 2002). Oppgaver som passer med læringsmålene og læringsutfallene er skissert i delkapittel 3.3.1 – 3.3.4.

1. Definering av læringsmål for oppgavene	• Lære å identifisere de fire representasjonsformene av lineære funksjoner (A, N, G & S). • Lære å oversette lineære funksjoner fra algebraisk-, tabellmessig- og situasjonsrepresentasjon til grafisk representasjon. • Forstå begrepene *stigningstall* og *konstantledd/skjæringspunkt med y-akse* og vite hvordan disse kritiske egenskapene ved lineære funksjoner påvirker den aktuelle grafens utseende og verdier.
2. Spesifisering av konkrete læringsutfall	• Kapabilitet til å gjennomføre oversetting av numeriske tabeller, situasjonsbeskrivelser og algebraiske uttrykk som representerer lineære funksjoner til grafisk representasjon.
3. Konseptuelle og prosedurale kunnskaper som kreves av oppgavene	• Plotting av punkter krever at eleven kan overføre mengder av ordnede par til koordinatsystemet • Skissering av funksjonsuttrykk og modellering av situasjonsbeskrivelse

	• Kjennskap til de fire representasjonsformene til en funksjon • Kunnskap om begrepene *stigningstall* og *konstantledd/skjæringspunkt med y-akse*

Tabell 2, læringsmål, læringsutfall, samt de konseptuelle og prosedurale ferdighetene som kreves av oppgavene (egen produksjon).

I tabell 2 over vises læringsmålene, de potensielle læringsutfallene, samt de konseptuelle og prosedurale kunnskapene som kreves i oppgavene. Det siste læringsmålet i tabellen er spesielt viktig, fordi det er disse egenskapene ved lineære funksjoner som definerer en grafs utseende. Det er ikke eksplisitt nødvendig å kjenne til konseptene stigningstall og konstantledd for å kunne plotte punkter fra en numerisk tabell, og trekke en linje mellom dem. Likevel fungerer læringsmålet som en såpass fremtredende faktor for å lykkes med det ønskede læringsutfallet at det ikke kan oversees.

3.3 Design av oppgaver i testen

Ettersom strukturen i eksperimentet og testen er presentert i delkapittel 3.2.1 og 3.2.2 vil jeg i dette kapitlet forklare hvordan oppgavene i testen er konstruert. Oppgavenes innhold er definert etter læringsmålene og læringsutfallet i dem.

På bakgrunn av teori presentert i kapitel 2.2 ble en test bestående av fem digitale representasjonsoppgaver designet. De to første omhandler oversettelse fra numerisk tabell (N) til graf (G) og den tredje om oversettelse fra algebraisk likning (A) til graf. De to siste oppgavene ber brukeren oversette en situasjonsbeskrivelse (S) til graf. Oppgavene ble utformet med tanke på økende kompleksitet gjennom oppgavesettet etter Adu-Gyamfi et al. (2014; 2012) og Bossé et al (2011). Her er den antatt minst krevende oppgaven den første, og den antatt mest krevende oppgaven ble gitt til sist.

Med hensyn på de kontaktede matematikklærernes respons og den nåværende læreplanens kompetansemål i funksjoner, ble oversettelsesoppgavene utformet til å handle om lineære funksjoner. Alle oppgavene i testen er like i den forstand av at elevene besvarer dem gjennom å manøvrere seg i grafikkfeltet i programvaren Geogebra. Elevene skal manipulere to punkter A og B slik at en lineær graf som er tegnet gjennom disse punktene korrekt representerer den numeriske tabellen, det algebraiske uttrykket eller situasjonsbeskrivelsen som gis i den aktuelle oppgaven.

Generelt vil samtlige oppgaver ha noen feiltilstander som programvaren kan reagere på. Disse kan eksempelvis være å finne galt fortegn på stigning, men korrekt konstantledd, finne korrekt fortegn på stigningen, men galt stigningstall og korrekt konstantledd og så videre. Alle kombinasjoner av å ha helt galt, delvis korrekt og korrekt representert grafen gjennom stigningstall og konstantledd må tas høyde for i utformingen av oppgavene. I tillegg finnes det flere spesifikke feil som kan oppstå for hver enkelt oppgave. Disse spesielle feilene vil jeg diskutere i delkapitlene under, sammen med de ulike oppgavenes generelle utforming og vanskegrad gitt forskningen til Adu-Gyamfi et al. (2014; 2012) og Bossé et al. (2011).

3.3.1 Oppgave 1 og 2

Oppgavenes vanskelighetsgrad, utforming, typiske feil og feilkilder beskrives ut ifra:
Oppgave 1: lav attributt-tetthet, lite forvirrende fakta, lite faktagap og lokal oversettelseshandling (plotting)
*Oppgave 2: lavere attributt-tetthet, middels forvirrende fakta, lite faktagap og lokal oversettelseshandling (plotting) (*Adu-Gyamfi et al., 2014; 2012; Bossé et al., 2011).

x	y
0	5
1	0
2	-5
3	-10
4	-15
5	-20

x	y
-6	-27.5
-4	-19.5
-2	-11.5
0	-3.5
2	4.5
4	12.5
6	20.5

Figur 8, figur av tabellene elevene fikk presentert i oppgave 1 og 2

De to første oppgavene (se appendiks nr. 1 og 2), ber elevene tilpasse grafen de ser i grafikkfeltet slik at den passer med tabellene over, noe som krever en lokal oversettelseshandling – plotting. Den første tabellen inneholder x-verdier som inkrementer med økning på 1. Dersom en ser til x-verdiene, starter også på 0, og sammen med en svært synlig stigning i y-verdi på -5 var det forventet at de aller fleste elevene kunne plotte denne numeriske tabellen.

Oppgave 2 sin medhørende tabell er noe mer kompleks i den forstand av at x-verdiene nå starter på -6, samt at det inkrementerer med 2 for hvert ordnede par. Dette maskerer funksjonens stigningstall noe mer og gir dermed tabellen litt lavere attributt-tetthet enn første tabelloppgave. I tillegg inneholder y-verdiene flere ordnede par, og inkluderer større og lavere desimaltall, noe som gir den middels mengde forvirrende fakta. Det var dermed forventet at oppgave 2 er den som nest flest elever ville få til.

For å mestre disse oppgavene kreves det kun at elevene kan overføre minst to ordnede par til koordinatsystemet slik at linjen mellom dem matcher med tabellen. På det grunnlaget vil tolkningsfeil antagelig oppstå hyppigst (Adu-Gyamfi et al., 2012). Presumptivt vil noen elever forsøke å flytte på punktene A og B i koordinatsystemet slik at f.eks. punkt A blir (5,0), der det skulle vært (0,5) (Bossé et al, 2011).

I tillegg kan syntaktiske feil eller tolkningsfeil dukke opp som følge av en generell mangel på forståelse av hva tabellene egentlig representerer relativt til en graf. Enkelte elever vil kanskje lese av tabellene nedenfra og opp slik at det kan virke som at grafens stigning vil omvendes, altså en form for manipulasjonsfeil.

3.3.2 Oppgave 3

*Oppgave 3: høy attributt-tetthet, lite forvirrende fakta, lite faktagap og global oversettelseshandling (skissering) (*Adu-Gyamfi et al., 2014; 2012; Bossé et al., 2011).
Den tredje oppgaven (se appendiks nr. 3), ber elevene tilpasse grafen i grafikkfeltet slik at den passer med uttrykket $y = -3x + 6$. Verdiene som opptrer som stigningstall og konstantledd er i dette tilfellet relativt enkle å forholde seg til – uten desimaler, svært høye eller lave tallverdier. Uttrykket inneholder ingen ekstra ledd eller forvirrende fakta, slik at attributt-tettheten er svært høy. Kun den relevante informasjonen presenteres i uttrykket, slik at selve oversettelseshandlingen skissering er det som gjør denne oppgaven presumptivt vanskeligere enn de to foregående.

I motsetning til forrige oppgave er elevene her absolutt nødt til å forholde seg til begrepene stigningstall og konstantledd. De må altså forstå at i et uttrykk på formen $y = ax + b$, vil a referere til hvor mye grafens y- verdi endrer seg per x- verdi. Det var på forhånd forventet at en tolkningsfeil i form av å bokstavelig tolke uttrykket til at grafen bør gå gjennom punktene $(-3,0)$ og $(0,6)$ ville oppstå blant enkelte elever (Klegseth, 2018). For andre elever kan det være nødvendig å beregne verdiene i uttrykket for å så plotte punktene slik at implementeringsfeil også kan oppstå underveis i utregningen.

Manipulasjonsfeil kan også forekomme gjennom å blande mellom stigningstallet og konstantleddets rolle slik at grafen får en stigning på 6. Generelt vil også delvise feil kunne oppstå. Tilfeller der eleven har funnet rett stigningstall, men kanskje tror at grafen skal skjære x-aksen ved $(6, 0)$, eller andre kombinasjoner av å nærme seg rett svar kan finne sted. I tillegg kan eleven finne på å konstruere en graf uten stigning eller lage andre tilfeller av grafer som ikke nærmer seg nærmer seg rett stigning eller skjæring med y-aksen.

3.3.3 Oppgave 4

*Oppgave 4: middels attributt-tetthet, middels forvirrende fakta, lite faktagap og global oversettelseshandling (skissering) (*Adu-Gyamfi et al., 2014; 2012; Bossé et al., 2011).
I denne oppgaven (se appendiks nr. 4) blir elevene forklart følgende: «Du sitter i en bil ved et lyskryss. I det trafikklysene skifter til grønt setter du foten på gasspedalen. Bilen øker sin hastighet med $10\ \mathrm{km/t}$ hvert andre sekund du sitter i bilen. Dra i punktene A og B slik at grafen som tegnes mellom dem viser bilens hastighet fra trafikklyset blir grønt.».

Innføringen av en skriftlig oppgave som spesifikt nevner bilens økende hastighet krever på en side forståelse av konseptet stigningstall, men situasjonens virkelighetsnærhet kan for enkelte elever gjøre den mer intuitiv. Den ønskede grafens skjæringspunkt med y-aksen blir heller ikke beskrevet eksplisitt, men ligger noe maskert i siste setning av oppgaveteksten. Her er det ønskelig at elevene skal oppdage at bilens stillestående utgangspunkt ved tid lik null gjør at grafen skjærer gjennom koordinatsystemets origo for å deretter å øke med $5\ \mathrm{km/t}$ per sekund. Elevene kan gjøre en feil hvorpå de setter grafens stigningstall til å være nettopp 10, noe som kan klassifiseres som en syntaktisk-strategi feil.

I denne oppgaven har jeg endret de to aksene i koordinatsystemet til å eksplisitt vise hastighet i kilometer per time og tid, men likevel forventer jeg at enkelte elever blander mellom disse. En del elever kan også konstruere en graf uten stigning, da det kan virke intuitivt for noen å tenke seg at bilens hastighet endrer seg lineært, og dermed bør grafen også være «flat» (Ainsworth et al., 2002; Bossé et al., 2011).

3.3.4 Oppgave 5

*Oppgave 5: middels attributt-tetthet, middels forvirrende fakta, lite faktagap og global oversettelseshandling (skissering) (*Adu-Gyamfi et al., 2014; 2012; Bossé et al., 2011).

I oppgave 5 (se appendiks nr. 5) blir elevene forklart følgende: «Tenk deg at du vil ha en plante på eiendommen din. Tidlig på våren planter du et lite tre på 24 centimeter i hagen. Hvert år etter treet ble plantet registrerer du at det har vokst 14 centimeter. Tilpass grafen med punktene A og B slik at den viser hvor høyt treet ditt er etter et gitt antall år.».

Her er det ønskelig at elevene skal oppdage at treets høyde ved planting tilsvarer grafens skjæringspunkt med y-aksen og vekstraten på 14 centimeter skal tilsvare grafens stigningstall. Ettersom jeg har skrevet opp konstantleddet først var det dermed forventet at noen elever ville få feil knyttet til syntaktisk strategi. Enkelte elever vil også kunne få implementeringsfeil knyttet til å legge sammen tallene 24 og 14 for å konstruere grafen ut ifra treets første vekstår. I tillegg vil overordnede manipulasjonsfeil som forvirring rundt stigningstallets fortegn forekomme.

Delvise feil, for eksempel tilfeller der eleven finner korrekt konstantledd, men galt stigningstall kan i likhet med de øvrige oppgavene også komme til syne. Det vil i tillegg være verdt å merke seg at i likhet med oppgave 3 og 4 (se appendiks 3 og 4) er oversettelseshandlingen i å gå fra en situasjonsbeskrivelse til grafisk representasjon en global handling (skissering), noe som betyr at elevene må konseptuelt forholde seg til to samvarierende variabler.

Generelt bør det nevnes at av de mange egenskapene digitale verktøy som Geogebra bringer med seg uteblir i oppgavesettet jeg har designet. I kapittel 2.2.2 skisserer jeg aspektene, multiple representasjoner, dynamiske egenskaper, kalkuleringskapabilitet og umiddelbar tilbakemelding. Av disse er det de dynamiske egenskapene, den grafiske visualiseringen innenfor kalkuleringskapabilitet og tilbakemelding som gjenstår i programvaren jeg har benyttet meg av. Om for eksempel det algebraiske uttrykket til elevenes konstruerte graf hadde blitt vist samtidig som de endret grafen, ville det i praksis avslørt oppgavens korrekte løsning. I sammenhengen av å teste elevenes oversettelsesferdigheter mot grafisk representasjon ville det vært det vært ugunstig, men i en realistisk undervisningssituasjon kunne spesielt flerfoldig visning av representasjoner vært nyttig.

3.4 Design av tilbakemeldinger

I dette delkapittelet besvares første delspørsmål i problemstillingen: «Hvordan kan det designes automatiske formative tilbakemeldinger til oppgaver om oversettelse fra tabellform, algebraisk- og situasjonsrepresentasjon til grafisk representasjon av lineære funksjoner?». Denne undersøkelsens valg av innhold i tilbakemeldingene er bare én av mange mulige, men tegner likevel et bilde av *hvordan* det kan gjøres ved hjelp av bug-related tutoring (feilrelatert veiledning) (Narciss & Huth, 2002). I første omgang vil jeg beskrive de utdypende og verifiserende tilbakemeldingene eksperimentgruppen fikk. Deretter forklarer jeg de rent verifiserende tilbakemeldingene kontrollgruppen ble vist.

3.4.1 Tilbakemeldinger til eksperimentgruppen

Tilbakemeldingene til den eksperimentelle gruppen er i likhet med prinsippene for tilbakemelding i artiklene til Narciss & Huth (2002) og Narciss (2004; 2013) designet for å gi tilbakemeldinger som blir gradvis mer hjelpende dersom eleven foretar samme type feil gjentatte ganger. I tillegg er enkelte av tilbakemeldingene noe preget av Mevarech & Kramarski (1997, 2003) og Kramarski & Zeichner (2006) sine ideer om metakognitive motspørsmål. Når jeg bruker begrepet «samme type feil» gjenspeiler det hvilken feiltilstand programvaren reagerer på.

Ettersom elevene jobber innenfor et koordinatsystem undersøker programvaren hvorvidt elevens funksjon har nådd svaret ved å se på dets stigningstall, konstantledd og hvilke koordinater de to punktene som ligger på funksjonslinjen er plassert på. Enkelte elevsvar kan derfor være nærmere et korrekt svar. Eksempelvis kan grafens skjæringspunkt med y-aksen være korrekt plassert, men dens stigning kanskje er negativ fremfor positiv.

Tilbakemeldingene eksperimentgruppen for inneholder en verifiserende og utdypende del. Med verifiserende menes en beskjed om i hvilken grad elevens svar er korrekt, mens den utdypende delen dreier seg om et hint om strategier mot riktig løsning. Det kan være hint om feilen eleven har gjort, hint til en metakognitiv strategi, hint til relevant konseptuell kunnskap eller prosedural ferdighet, metakognitivt motspørsmål relativt til elevresponsen, forklaring på en begått systematisk feil eller hint til relevante begrepers betydning. Hintene blir mer tydelige og spesifikke jo fler feil av samme kategori en elev gjør. På oppgave 3 ($y = -3x + 6$) (se appendiks 3) kan en elev konsekvent finne riktig konstantledd i 6, men ender opp med å ha positiv stigning. Dersom eleven begår denne typen feil tre ganger, vil programmet gi beskjedene:

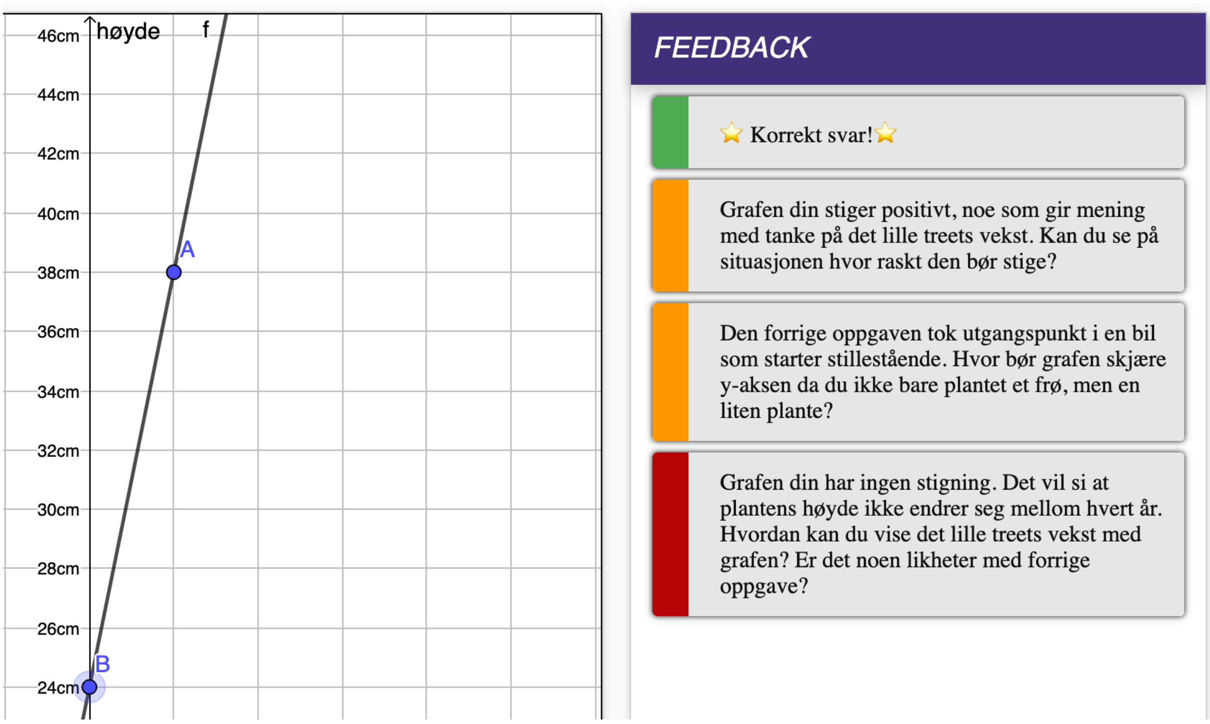
1. "Du har funnet rett skjæringspunkt med y-aksen, men du må finne grafens stigningstall. Hvordan kan du se på uttrykket hva stigningstallet er?",

2. "Ettersom grafen din skjærer y-aksen i $(0, 6)$ er dette riktig, men du mangler å finne rett stigning på grafen. Stigningstallet står i uttrykket som -3.",

3. "Når stigningstallet er -3 betyr det at funksjonen synker med 3 y-verdier for hver økning i x-verdi."

Tilbakemeldingene som eleven får presentert vil i første rekke presenteres med et motspørsmål. Deretter vil neste tilbakemelding presisere hva stigningstallet er i funksjonens uttrykk før den spesifiserer hva dette stigningstallet betyr. Den siste tilbakemeldingen er i eksempelet som er gitt over, veldig tydelig, men den gis også ved elevens tredje gjentagende feil. Alle tilbakemeldingene som gis gjennom programmet er konstruert etter en tilsvarende logikk. De blir gradvis mer styrende mot et feilelements betydning for grafens utseende.

I tillegg presenteres de med en fargekode. Tilbakemeldingene over ville blitt gitt med en gul farge, mens en mer graverende feil som å tolke uttrykket $y = -3x + 6$ bokstavelig som $y = 2x + 6$ ved at den derfor går gjennom punktet (-3, 0) og (0, 6) presenteres med en rød farge slik en kan se i figur 9.



Figur 9, utdypende tilbakemeldinger gitt som respons på fire ulike besvarelser på oppgave 5 (egen produksjon).

Tilbakemeldingene gis i det øyeblikket eleven trykker på en «sjekk svaret»- knapp, og deretter stables de vertikalt slik at den siste tilbakemeldingen ligger øverst. Dette er en designmessig vurdering som muligens kan kritiseres. Ved å vise alle de gitte tilbakemeldingene kan det med mange forsøk bli forvirrende for eleven som arbeider å vite eksakt hvilken tilbakemelding som peker til en tidligere svartilstand. Samtidig ble avgjørelsen om å holde på de gitte tilbakemeldingene tatt for at elevene kunne se tilbake på hva programmet har gitt av informasjon.

Det er ikke bare oppgave 3 med *A*→*G* oversettelse som inneholder tilbakemeldinger med de kritiske begrepene stigningstall og konstantledd/skjæringspunkt med y-aksen. Samtlige av oppgavenes tilbakemeldinger fokuserer på disse konseptuelle elementene. Det å forstå lineære funksjoner handler om mer enn bare å anvende prosedyrer, men involverer det å forstå sammenhengen mellom de ulike representasjonene funksjonene inntar (Adu-Gyamfi & Bossé, 2014; Chiu et al., 2010). Funksjonenes stigningstall og konstantledd vil altså være definerende for dets utseende, noe som gjør at disse egenskapene vektlegges i tilbakemeldingene, uansett oppgave.

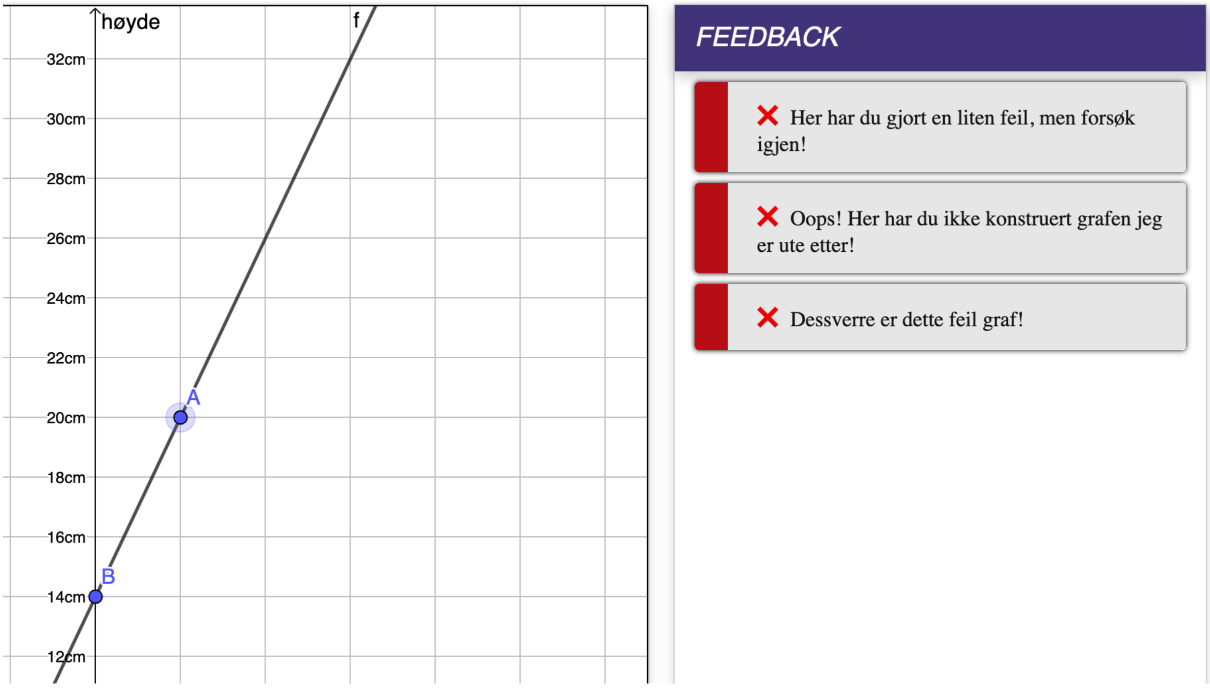
Tilbakemeldingene i oppgave 4 og 5 med S→G oversettelse, (se appendiks 4 og 5) skiller seg dog noe ut ved at de forsøker å knytte de kritiske begrepene til konteksten i oppgaven. Når en bil akselerer slik som i oppgave 4, kan bilens fartsøkning knyttes til grafens stigningstall. Utgangshastigheten bilen har vil dermed kunne assosieres til grafens skjæringspunkt med y-aksen. På samme vis vil spirens årlige vekst i oppgave 5 kunne beskrives som den tilknyttede grafens stigningstall, samtidig som høyden den har ved planting forbindes med grafens skjæringspunkt med y-aksen. Jeg antok at de kritiske begrepenes tilstedeværelse i samtlige oppgavers tilbakemelding, ville føre til at den eksperimentelle gruppens prestasjoner kom til å skille seg mer fra kontrollgruppen jo lenger ut i oppgavesettet elevene kommer.

Samtidig er det dermed en fare for at tilbakemeldingene i oppgave 1 og 2 med T→G oversettelse, (se appendiks 1 og 2) kunne virke noe forvirrende for elevene. Oversettelse fra numerisk tabell til graf regnes som en lokal oversettelseshandling – plotting (Adu-Gyamfi et al., 2012; Bossé et al., 2011). I en slik situasjon kreves det i realiteten kun at eleven kan konstruere en graf fra et sett med ordnede par. Altså vil forståelse av de kritiske begrepene i disse oppgavene være overflødig for oppgave 1 og 2.

Jeg prioriterte læringsutfallet for oppgavesettet som helhet, hvorpå stigningstallet og konstantleddet er definerende for grafenes utseende. Tanken var at elevene skulle se sammenhengen på tvers av oppgavene. Goldenberg påpeker at en sammenhengende, tingliggjort forståelse av funksjoner ikke oppstår i isolert arbeid med enkeltrepresentasjoner: «*Translation across multiple representations can help reduce the isolation of each mathematical lesson and help provide a coherent and unified view of mathematical method and content.*» (Goldenberg, 1988 i Adu-Gyamfi et al., 2012, s. 159).

3.4.2 Tilbakemeldinger til kontrollgruppen

Tilbakemeldingene som kontrollgruppen får vil kun være av verifiserende karakter. Verifiseringen er i tillegg mer begrenset med tanke på informasjonen den gir. Den forteller eleven om det avgitte svaret er korrekt eller ei, men ikke i hvilken grad eller hvor feilen ligger. I tillegg gis beskjeder om feil med en rød fargekode uavhengig av hvor nære eleven er å få til oppgaven. Tilbakemeldingene er altså for kontrollgruppen rent dikotomiske, og dermed til liten hjelp utover å informere om svaret er riktig eller galt, slik en kan se i figur 10 under.



Figur 10, verifiserende tilbakemeldinger gitt som respons på tre ulike elevsvar på oppgave 5 (egen produksjon)

Generelt vil tilbakemeldingene kontrollgruppen får servert i praksis være lik den Maple T.A. gir. Programmet reagerer kun statisk på hvorvidt oppgaven er løst korrekt eller ei, og fungerer slik sett som en motvekt til de feilrelaterte tilbakemeldingene eksperimentgruppen får (Narciss, 2013).

3.4.3 Antallet tilbakemeldinger

Tilbakemeldingsmetodikken jeg har skissert i kapittel 3.4 ligner altså ikke nevneverdig på noen av modellene som dagens private aktører bruker, med et mulig unntak av Gyldendal Smartøving. Khan-Academy og Maple T.A. benytter seg av et statisk system som gir samme tilbakemelding i form av fasit på oppgavene og løsningsstrategi. Kikora gjør oppgaven for brukeren når en trykker på hint-knappen i systemet deres.

Systemet jeg har satt sammen og bruker, etterstreber å respondere på brukerens handlinger relativt til deres svarforsøk. Samtidig har det en utfordring som gjelder mengden av innhold som må ligger i ulike svartilstander. Elevene som tar testen har 20 forsøk på hver oppgave, og kan dermed teoretisk sett gjøre samme type feil 20 ganger. Tilbakemeldingssystemet mitt inneholder ikke 20 forskjellige typer tilbakemeldinger for hver enkelt feil, men opp til fem. Skulle det gjort det, måtte det blitt konstruert $20 \times 8 \times 5 = 800$ unike tilbakemeldingsstrenger for kun fem ulike oppgaver. Dette aspektet illustrerer et problem ved å benytte teori til å forsøksvis forutsi hvilke feil elevene kan gjøre, for å så konstruere tilbakemeldinger for disse feilene. Det må genereres svært mange tilbakemeldinger (Rønning, 2017).

Når en elev eventuelt har utvist samme type feil så mange ganger at alle tilbakemeldingsstrengene for én type feil er gitt, gir programvaren beskjed om at den ikke finner noe nytt å gi tilbakemelding på.

3.5 Erfaringer fra pilotundersøkelse

I forkant av den reelle datainnsamlingen måtte det undersøkes om oppgavene og testen som helhet fungerte som tiltenkt i Matistikk-plattformen. Aller først fikk fem medstudenter prøve seg på oppgavesettet. Deres gjennomføringer og tilbakemeldinger avslørte noen små, men likevel betydningsfulle elementer som ble modifisert.

Eksempelvis hadde enkelte problemer med å få punktene de skulle manipulere til å «feste» seg på det tiltenkte koordinatet. Datamaskinens vurdering av svarene krever presisjon. Dersom brukeren ønsker å plassere punkt A i koordinatet $(5, 2)$, men ender opp med å legge det i eks. $(5.001, 2.004)$, vil du risikere å få galt svar selv om det ved første øyekast kan se ut til å være riktig. Løsningen på problemet var å programmere Geogebra til å automatisk bedømme hvilken hel eller halv inkrementering av koordinater punktet ligger nærmest. Hensikten med denne endringen var forsøksvis å eliminere elevfeil som ikke kommer som følge av manglende forståelse.

I etterkant av denne mindre testingen på medstudenter ble en pilotundersøkelse gjennomført ved en 10. klasse i Trondheimsregionen. Formålet i denne omgangen var å sikre seg at det ikke oppstod tekniske problemer og at oppgavetekstene var tydelig nok formulert. Elevene ble bedt om å skyve pultene i god avstand fra hverandre, samt unngå kommunikasjon seg imellom. Jeg forklarte at hverken meg selv, lærer eller medelever kunne hjelpe hverandre med oppgavene. Dersom de hadde spørsmål om oppgavenes løsning ble de bedt om å lese oppgaveteksten én gang til, samt iaktta tilbakemeldingene i programvaren. Elevene i klassen ble deretter delt inn i eksperimentgruppen og kontrollgruppen.

3.6 Gjennomføring av undersøkelsens datainnsamling

Selve gjennomføringen av undersøkelsen ble i praksis gjennomført likt på de ulike skolene slik den hadde blitt utført i pilotundersøkelsen. Da jeg ikke endret oppgavene eller tilbakemeldingene som ble gitt i dem, kunne dermed data fra pilotundersøkelsen inkluderes i det øvrige datamaterialet. Etter pilotundersøkelsen ble fire nye klasser fra tre ulike skoler inkludert for å gjennomføre testen. Testene ble gjennomført i løpet av en periode på tre uker i november. Ved hver gjennomføring av testen i en klasse, benyttet jeg 15 minutter til å fortelle om testen, dets innhold, hvordan manøvrere seg i menyene og grafikkfeltet, og at datamaterialet ble samlet inn til en masteroppgave. På denne tiden ble også elevene i hver klasse tilfeldig inndelt i de to testgruppene, slik det beskrives i kapitlet om undersøkelsens validitet, kap. 3.8.1. Hver klasse hadde totalt 45 minutter til rådighet for å gjennomføre testen.

En utfordring jeg støtte på under datainnsamlingen, men ikke oppstod under pilotundersøkelsen, var tekniske problemer av to typer. Den første var et tilfelle der internettilgangen til elevene forsvant mot slutten av testøkten ved en skole. Dette er et element som er uheldig i sammenheng med digital datainnsamling, men ettersom dette inntraff mot slutten av testøkten, rammet det kun fem elever.

Det andre tilfellet av tekniske problemer inntraff som følge av at en klasse hadde svært mange elever. Dette gjorde at fire elever ble kastet ut av testen mens de var halvveis til å gjennomføre, mens noen brukte lang tid på å komme seg inn på nettsiden. Antageligvis kommer dette av begrenset serverkapasitet i Matistikk-plattformen.

3.7 Scoring av oppgaver og valg av analysemodell

Rådatamaterialet jeg sitter på er altså en output fra programvaren i testen jeg har brukt i form av JSON-strenger lagret i csv-format. Det vil si tekststrenger fra hver enkelt bruker, som inneholder informasjon om hvordan de har plassert sin graf, de to punktene som definerer grafen mellom seg, hvor lang tid de har brukt, om det endelige svaret er korrekt eller ei og hvilken tilbakemelding de har fått underveis. En kan tenke på rådataene som en slags «historie» som kronologisk forteller om arbeidsprosessen til elevene.

I første omgang måtte datamaterialet forberedes for kvantitativ analyse med programmet IBM SPSS. Dette ble gjort ved å telle opp antall korrekte, ukorrekte og ufullstendige svar. De to første begrepene er selvforklarende, mens elevsvar som ble avgitt i et tidsrom på mindre enn 5 sekunder ble markert som ufullstendige. Ettersom enkelte elever forsøkte å laste inn testen på nytt, ble gjenværende spørsmål også markert som ufullstendige. Eksempelvis kan en elev ha gjennomført de tre første oppgavene før hun bestemte seg for å laste inn siden på nytt, noe som resulterte i at de to siste oppgavene ble markert som ufullstendige. Andre hendelser som at en elev ikke har rukket å gjøre ferdig alle oppgavene, eller at nettilgangen forvsinner, fører også med seg at de ubesvarte oppgavene regnes som ufullstendige.

De ufullstendige svarene ble satt opp i SPSS, men jeg instruerte programvaren til å regne de som «missing values» - data som ikke medgår i analysen. Dersom en elev eksempelvis har 3 godkjente svar og 2 ufullstendige svar, vil kun de godkjente svarene beregnes i analysen av enkeltoppgavene.

For å kunne bestemme hvilken effekt tilbakemeldingene fra programvaren hadde på den eksperimentelle gruppen versus kontrollgruppen ble det gjennom ført en uavhengig t-test på hver enkelt oppgave og på testen som helhet. En t-test egner seg når en vil avgjøre om det er statistisk signifikant forskjell mellom gjennomsnittene i to ulike eller like grupper hentet fra et tilfeldig, normalfordelt utvalg (Cohen et al., 2018; Hutchison & Styles, 2010; Robson, 2002). I de tilfeller en sammenligner to ulike grupper vil en uavhengig t-test gjøre seg gjeldende.

I denne undersøkelsen er to forskjellige grupper bestående av ulike individer plassert tilfeldig i henholdsvis kontrollgruppen eller den eksperimentelle gruppen – ergo kan en uavhengig t-test velges i dette tilfellet. T-testen vil brukes for å avgjøre om det finnes en signifikant forskjell i resultatene fra testen med tanke på elevenes prestasjoner og tidsbruk.

Jeg vil bemerke at jeg vurderer resultatene av t-testene med utgangspunkt i en to-halet test. Grunnen til å velge den varianten kommer av at jeg ikke kan utelukke en potensiell negativ effekt av de utdypende tilbakemeldingene på den eksperimentelle gruppen. Videre tar t-testen utgangspunkt i en hypotesetest. Dette gjør jeg ved å anta en nullhypotese, H_0 og en alternativ hypotese, H_1 som er gjeldende for alle de fem oppgavene, samt testen som helhet (Cohen et al., 2018):

> H_0: Det er ingen statistisk signifikant skårforskjell i gjennomsnittsresultatene mellom kontrollgruppen og den eksperimentelle gruppen ved $\alpha = 0.05$.
>
> H_1: Det er statistisk signifikant skårforskjell i gjennomsnittsresultatene i kontrollgruppen og den eksperimentelle gruppen ved $\alpha = 0.05$.

Dersom jeg ikke finner en signifikant skårforskjell i gjennomsnittsresultatet på tvers av de to gruppene i en oppgave vil det medføre at jeg støtter nullhypotesen. På samme vis vil en signifikant forskjell føre med seg at jeg støtter den alternative hypotesen, som antar at de utdypende tilbakemeldingene gitt til kontrollgruppen gir forskjell i gjennomsnittsskår sammenlignet med kontrollgruppen. Signifikansnivå på 5% er valgt på grunnlag av å være standarden mange vitenskapelige artikler belager seg på (Cohen et al., 2018). Det vil si at sannsynligheten for at det oppnådde resultatet kan oppstå ved en tilfeldighet må være lik eller lavere enn 5% for å kunne aksepteres. Når signifikansen for t-verdiene beregnes vil de fremkomme som p-verdi. Dersom p-verdien for en t-test eksempelvis er $p = .03$ ser jeg at denne verdien er lavere enn signifikansnivå på 0.05, noe som betyr at den alternative hypotesen vil støttes.

Samtidig må jeg erkjenne at statistisk signifikans er kalkulert som en funksjon av testutvalgets størrelse, noe som betyr at det ofte vil bli funnet med stort nok utvalg (Cohen et al., 2018; Robson, 2002). Dermed presenterer jeg også den kalkulerte effektstørrelsen av de utdypende tilbakemeldingene gjennom Hedges *g*. Fordelen med å akkompagnere beregnet effektstørrelse med t-testene er at effektstørrelsen vil fortelle meg *hvor stor* effekten eller differansen mellom kontrollgruppen og den eksperimentelle gruppen er. Dette vil ikke statistisk signifikans alene si noe om (Cohen et al., 2018; Ellis, 2019).

Generelt vil effektstørrelser beregnes som forskjellen i gjennomsnittsresultatene mellom de testede gruppene delt på standardavviket som utvalget er hentet fra (Ellis, 2019). Den vanskelige delen kan beskrives som å finne standardavviket i gruppen som utvalget er hentet fra, men her kan det sammenslåtte utvalgets standardavvik brukes. Det finnes noen populære varianter av effektstørrelser, som Cohens *d*, Glass' Δ og Hedges *g* (Ellis, 2019). Hedges *g* kalkuleres ut fra formelen $g = \frac{m_A - m_B}{SD^*_{sammenslått}}$, hvor m_A er gjennomsnittsskår i den eksperimentelle gruppen og m_B er gjennomsnittskår i kontrollgruppen, og $SD^*_{sammenslått} = \sqrt{\frac{(n_A - 1)SD_A^2 + n_B - 1)SD_B^2}{n_A - n_B - 2}}$ (Ellis, 2019). Her fremtrer n_A som størrelsen i den eksperimentelle gruppen, n_B som størrelsen i kontrollgruppen, SD_A som standardavviket i den eksperimentelle gruppen og SD_B som standardavviket i kontrollgruppen.

Beregningen gir effektstørrelser vektet etter de relative størrelsene i de to utvalgene (Ellis, 2019; Hedges, 1981). Ettersom kontrollgruppen og den eksperimentelle gruppen innehar noe forskjellige utvalgsstørrelser blir Hedges g benyttet i denne undersøkelsen, da den tar hensyn til diskrepans i størrelsen på utvalgene. $SD^{*}_{sammenslått}$ i Hedges g ble utviklet for å fjerne en liten positiv bias i favør av den eksperimentelle gruppen (Hedges, 1981). I praksis har denne korrigeringen liten betydning for resultatene som fremkommer i de fem delkapitlene som jeg vil presentere i analysedelen, men beregnes likevel fremfor den vanlige Cohens d for å oppnå størst mulig presisjonsgrad.

En effektstørrelse på $g = 1$ vil indikere at skårforskjellen mellom de to gruppene er lik ett standardavvik, $g = 2$ vil tilsvare to standardavvik og så videre. Cohens d og Hedges g tolkes på lik måte. Cohen antyder at en kan benytte seg av en slags «tommelfingerregel» til hjelp i tolkningen av effektstørrelser (Cohen J. , 1988; Cohen et al., 2018):

- Liten effekt (vanskelig å avsløre ved ren observasjon) $g = 0.2$
- Moderat effekt (effekten kan avsløres ved observasjon av trent personell) $g = 0.5$
- Sterk effekt (effekten er så stor at den er åpenbar ved førstehånds observasjon) $g = 0.8$

Jeg vil dog påpeke at å bruke begreper som «sterk» eller «liten» effekt kan være misvisende og ikke nødvendigvis til hjelp i å tolke et resultat. Hva som er en stor eller liten effekt på prestasjon er meningsløst uten et sammenligningsgrunnlag (Ellis, 2019). Derfor vil resultatene i denne undersøkelsen settes opp imot andre artiklers funn om automatiserte tilbakemeldinger i drøftingsdelen av denne undersøkelsen.

3.8 Forskningens troverdighet – validitet og reliabilitet ved oppgaven

For at et forskningsarbeid skal aksepteres innen et gitt fagfelt er det kritisk at forskningen utfører og beskytter krav om validitet og reliabilitet. Momenter som kan true en undersøkelses validitet og reliabilitet kan aldri helt fjernes, men minskes ved å ta bestemte grep (Cohen et al., 2018).

3.8.1 Validitet

Forskningslitteraturen skiller gjerne mellom intern og ekstern validitet i kvantitative undersøkelser og eksperimenter generelt. Den interne validiteten vil typisk ta stilling til om den eksperimentelle behandlingen faktisk utgjør en forskjell i effekten man ser på eksperimentgruppen (Cohen et al., 2018; Creswell, 2018).

For å sikre den ***interne validiteten*** i denne undersøkelsen har jeg tatt noen bestemte grep: randomisering av elever til de ulike testgruppene, testen som er benyttet ligger på en offentlig tilgjengelig server som videre gjør at studien kan repliseres, innsamlingen av data er forsøkt holdt lik på tvers av klassene som har deltatt i eksperimentet og anerkjente statistiske tester benyttes.

Når det gjelder randomiseringen av elevene til de to gruppene, burde det gjøres korrekt. Etter Hutchison & Styles (2010) forslag, ble elevenes navn listet opp i et databehandlingsprogram IBM SPSS, hvorpå hvert navn gis en tilfeldig verdi mellom 1 og 0 av en funksjon. Elever med allokerte verdier over 0,5 plasseres i kontrollgruppen, mens resterende plasseres i den eksperimentelle gruppen, slik at klassens deltagere og hvilken gruppe de havnet i ble randomisert algoritmisk. Samtidig var det ikke mulig for meg å fullstendig randomisere hvilke elever jeg fikk som utvalget til eksperimentet, slik jeg har skissert i kapittel 3.2.1.

I tillegg var det vanskelig å sikre paritet, altså likhet og sammenlignbarhet, mellom de to forsøksgruppene. Av tidsbegrensing hadde jeg ikke mulighet til å gjennomføre en pre- og posttest på gruppene, men kun én enkelt test. Denne faktoren er muligens den sterkeste trusselen mot eksperimentets validitet generelt.

Jeg har forsøkt å holde gjennomføringen av testen så lik som mulig på tvers av datainnsamlingsrundene på de ulike skolene. Til tross for dette finnes det faktorer som kan kontaminere resultatet av testen, men disse kan begrenses ved å ta noen valg. Eksempelvis trenger ikke elevene som gjennomfører testen bli informert om hvorvidt de er med noen gruppe overhode. Elevene ble derfor ikke fortalt at de ble delt inn i noen bestemte grupper, men ble kun bedt om å lese tilbakemeldingene de fikk nøye, uavhengig av gruppen de ble plassert i.

Kontaminering kan også oppstå som følge av kommunikasjon (Cohen et al., 2018). Lærere og elever kunne ikke kommunisere med hverandre gjennom testingen, og all informasjon som ble gitt til elevene var konsistent på tvers av datainnsamlingsrundene.

Ekstern validitet i kvantitativ utdanningsforskning omhandler hvorvidt en effekt en ser i et eksperiment kan generaliseres stabilt til den øvrige populasjonen som testutvalget er hentet fra (Cohen et al., 2018; Guba, 1981). Mennesker er komplekse skapninger med atferd som ikke nødvendigvis kan reduseres til snevre enkeltfaktorer. En bestemt endring i testbetingelser kan produsere et spesifikt resultat i en testrunde, men ikke i en annen. Spørsmålet om generaliserbarhet er dermed essensielt for å vurdere om et eksperiment faktisk har en betydning for personer utover testutvalget. For å sikre størst mulig generaliserbarhet må resultatene i det minste være statistisk signifikante, og utvalget bør være så stort som mulig. Testutvalget i denne undersøkelsen er på $n = 123$ elever, et ikke så veldig stort utvalg i kvantitativ sammenheng. Jeg må med andre ord utvise stor forsiktighet i å generalisere resultatene til den øvrige populasjonen av 10. klasseelever.

3.8.2 Reliabilitet

Generelt knyttes reliabilitet til presisjonen av undersøkelsens datamateriale – hvilke datatyper som brukes, viset de er innsamlet på, og hvordan datamaterialet til slutt er bearbeidet på. Hva gjelder reliabilitet i kvantitativ forskning deler Cohen, Manion & Morrison (2018) dette paraplybegrepet som handler om forskningens pålitelighet og etterprøvbarhet i to distinkte typer som hver for seg markerer reliabilitet. Disse er stabilitet og ekvivalens.

Stabilitet refererer til et mål på konsistens over tid. Det dreier seg altså om undersøkelsens evne til å produsere pålitelig stabile resultater over tid, gitt at utgangsbetingelsene ikke endres (Cohen et al., 2018; Creswell, 2018; Guba, 1981). For denne undersøkelsen vil det bety at dersom eksperimentet for eksempel hadde blitt gjennomført om et år på et utvalg av tiendeklasseelever, burde resultatet vært mer eller mindre likt. I praksis blir dette vanskelig for undertegnede å teste, og jeg tar dermed forbehold om stabilitetsfaktoren i denne sammenheng. På den annen side ble eksperimentet utført på flere skoler, i ulike klasser. Om samme tendens med tanke prestasjoner kan observeres for de ulike klassene, vil det være et argument i favør av reliabilitet i denne undersøkelsen.

Reliabilitet som ***ekvivalens*** vil ifølge Cohen et al (2018), Creswell (2018) handle om hvorvidt undersøkelsen kan produsere samme resultat gjennom replikasjon med ekvivalente former for datainnsamling. I praksis betyr det at et lignende eller alternativt måleinstrument eller test som fremstiller lignende resultater som de originale instrumentene, vil kunne sies å utvise reliabilitet. En pretest og deretter gjennomført posttest som gir lignende data, eller en statistisk test som viser de samme tendensene på flere grupper i utvalget kan også gi grunnlag for å hevde ekvivalens i undersøkelsen. Som nevnt tidligere er utelatelsen av pretest denne undersøkelsens kanskje største utfordring for dets validitet og reliabilitet.

3.9 Etiske betraktninger

Testen som elevene tok ble utført på NTNU sin Matistikk-plattform, hvorpå elevenes brukere blir identifisert som et firesifret tall. Det er derfor ingen måte jeg kan få innsikt i hvilken elev som tilsvarer en spesifikk besvarelse, da selv ikke IP-adressene til de som tar testen blir lagret. Resultatet av dette er at testen er fullstendig anonym. Det ble også presisert for elevene at det var frivillig å gjennomføre testen, men ettersom den ble tatt i en av deres matematikkundervisningsøkter kan det antageligvis ha virket som den var obligatorisk.

Det kan også virke urettferdig for enkeltelever som ble plassert i kontrollgruppen. Tilbakemeldingssystemet i denne undersøkelsen er tross alt designet etter beste hensikt om å promotere mestringsevne og læring, og når enkelte av elevene ikke får samme potensielle mulighet til dette kan det virke kritikkverdig (Hutchison & Styles, 2010). På den annen side er dette prosjektet gjennomført på bakgrunn av å forstå hvordan automatiserte tilbakemeldinger virker på elevene. Metoden kan dermed sies å tjene formålet, selv om det kan resultere i at noen enkeltelever i kontrollgruppen presterer svakere enn sine motparter i den eksperimentelle gruppen.

3.10 Metodekritikk

Når det gjelder tidspunkt for testingen, så forekom det i to av klassene at enkeltelever kom noe senere inn i klasserommene sine. Noen av dem hadde hatt samtaler med lærere i andre fag enn matematikk, eller kom av andre grunner litt forsinket inn til testtidspunktet. Disse elevene ble konsekvent plassert i den eksperimentelle gruppen, som er grunnen til at du som leser vil se at denne gruppen inneholder sju flere elever enn kontrollgruppen totalt sett.

Hva angår siste undervisningstidspunkt for funksjoner hadde samtlige klasser sist arbeidet med dette temaet mot slutten av foregående semester, altså rett før sommerferien. Det er ikke urimelig å anta at dette har en stor effekt på resultatet, spesielt med tanke på andelen korrekte svar per oppgave.

Det finnes også noen kritikkverdige punkter ved tilbakemeldingene i testen jeg designet. Programvarens bruk av begreper som stigningstall og skjæringspunkt med y-akse eller konstantledd kunne være til hinder for elevene. En elev kan eksempelvis ikke forstå begrepene stigningstall eller skjæringspunkt med y-aksen. Da vil ikke begrepene nødvendigvis være til hjelp i tilbakemeldingen. Disse begrepene ble også benyttet i oppgaver som ikke eksplisitt krever kjennskap til disse. I testens forsvar er noe av hensikten med programvaren å hjelpe elevene til å forstå og bruke disse begrepene i arbeid med oversettelse mot grafisk representasjon, uansett hvilken kilderepresentasjon som opptrer i oppgaven.

4 Resultater

I det påfølgende delkapittelet vil jeg vise undersøkelsens resultater. Her vil det andre delspørsmålet, «Hvordan påvirker automatiske verifiserende og utdypende tilbakemeldinger gitt av en datamaskin elevenes prestasjoner i oversettelsesoppgaver i lineære funksjoner?» besvares. Dette vil jeg gjøre gjennom å presentere resultatene fra hver enkelt oppgave med en uavhengig t-test, samt ved å vise effektstørrelsen av de utdypende tilbakemeldingene på den eksperimentelle gruppen gjennom Hedges g. Jeg vil også presentere en samlet analyse av forskjellen mellom den eksperimentelle gruppen og kontrollgruppens gjennomsnittskår. Videre vil en analyse av elevenes tidsbruk gjennom testen inkluderes i dette kapittelet.

Resultatene fra t-testene presenteres som $t(\text{frihetsgrader}) = t - \text{verdien}, p = p - \text{verdi fra 0 til 1}$. Deretter presenteres ($M = \text{gjennomsnittlig skår}, SD = \text{standardavvik}$), før effektstørrelsen kalkulert som Hedges g vises. Frihetsgradene viser antall uavhengige observasjoner fra hvert utvalg, og kalkuleres som antall observasjoner subtrahert med én fra hvert utvalg (Lysø, 2017). Eksempelvis vil en t-distribusjon med 117 frihetsgrader være gitt når et utvalg inneholder 63 observasjoner og det andre utvalget inneholder 56 fordi $(63 - 1) + (56 - 1) = 117$. Med økt mengde frihetsgrader minsker sannsynligheten for å få veldig lave eller høye t-verdier (Ellis, 2019).

T-verdien vil være avhengig av frihetsgradene fordi den kalkuleres relativt til populasjonsstørrelsene (Lysø, 2017). T-verdien måler differansen i skår mellom de to testpopulasjonene relativt til variasjonen i datamaterialet. Jo større t-verdi, dess større blir bevisbyrden mot nullhypotesen (Cohen et al, 2018; Ellis, 2019). Dersom t-verdien er negativ forteller det i vårt tilfelle om høyere gjennomsnittlig skår for kontrollgruppen relativt til eksperimentgruppen.

4.1 Oppgave 1 – tabell til graf

I oppgave 1 (se appendiks nr. 1), skal elevene representere en tabell som graf. Det var ingen signifikant gjennomsnittlig skårforskjell forskjell $t(117) = 0.489, p = .626$ på tvers av de 63 elevene i eksperimentelle gruppen ($M = 0.65, SD = 0.481$) og de 56 elevene kontrollgruppen ($M = 0.61, SD = 0.493$). Merk at førstnevnte gruppe produserte 4% større gjennomsnittlig skår for oppgave 1. Snittskårforskjellen i testskår mellom gruppene er svært liten og p-verdien fra den to-halede t-testen viser at nullhypotesen kan støttes for denne oppgaven. Ergo er det ingen beviselig forskjellig i snittskår for oppgaven. Samtidig kan jeg se til effektstørrelsen, ($g = .082$), som også forteller at effekten av intervensjonen den eksperimentelle gruppen ble utsatt for er svært liten, til ikke-eksisterende.

4.2 Oppgave 2 – tabell til graf

I oppgave 2 (se appendiks nr. 2), skal elevene representere en litt mer komplisert tabell som graf. Det var ingen signifikant gjennomsnittlig skårforskjell forskjell $t(108) = -0.489, p = .626$ på tvers av de 59 elevene i eksperimentelle gruppen ($M = 0.42, SD = 0.498$) og de 51 elevene kontrollgruppen ($M = 0.47, SD = 0.504$), selv om førstnevnte gruppe utviste en liten nedgang i gjennomsnittlig skår for oppgave 2. I motsetning til oppgave 1 finner jeg her at den eksperimentelle gruppen faktisk presterer noe lavere enn kontrollgruppen. I tillegg viser p-verdien fra den to-halede t-testen at nullhypotesen igjen bør støttes i for oppgave 2. For denne og foregående oppgave er det dermed ingen grunn til å tro at de utdypende maskingitte tilbakemeldingene har hatt en reell effekt. Dette gjenspeiler seg også i at effektstørrelsen er svakt negativ ($g = -0.10$).

4.3 Oppgave 3 – algebraisk uttrykk til graf

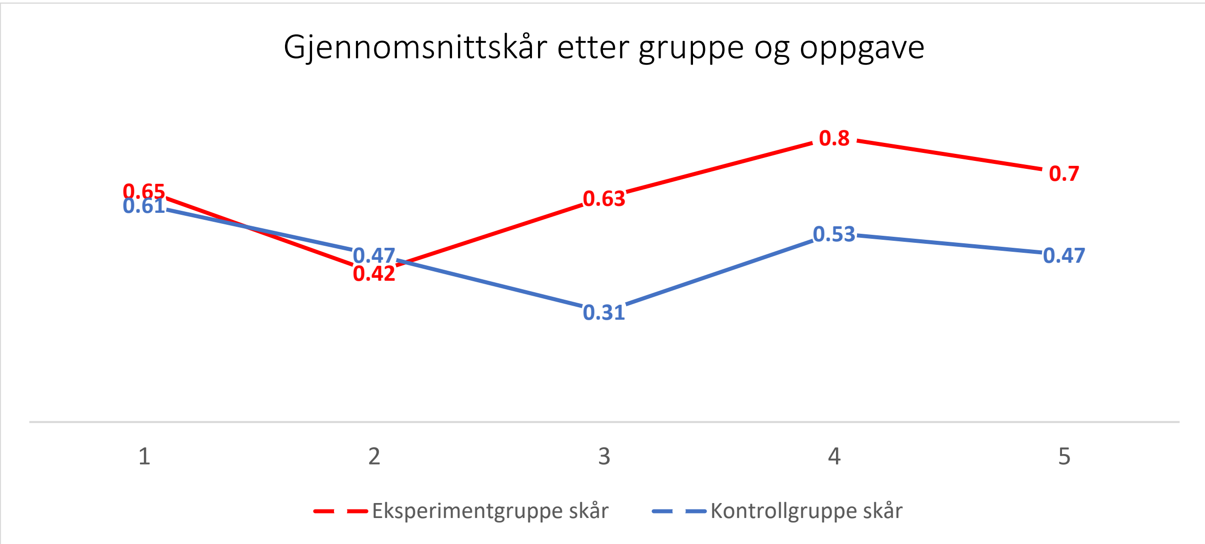
I oppgave 3 (se appendiks nr. 3), skal elevene representere funksjonsuttrykket $y = -3x + 6$ som en graf. De 57 elevene som ble utsatt for utdypende og verifiserende tilbakemeldinger som intervensjon ($M = 0.61, SD = 0.491$), sammenlignet med de 48 elevene i kontrollgruppen ($M = 0.31, SD = 0.468$) demonstrerte signifikant forskjellig gjennomsnittskår i oppgave 3, $t(103) = 3.201, p = .002$. For oppgave 3 finner jeg dermed grunn til å støtte den alternative hypotesen. Den eksperimentelle gruppen har hele 30% større snittskår på oppgave 3 enn det kontrollgruppen presterer. Dette er oppgaven hvor prestasjonsforskjellen er størst, noe som gjenspeiler seg i den moderate effektstørrelsen ($g = 0.62$). Da de utdypende tilbakemeldingene i tillegg til å fokusere på oppgaven har hovedfokus på lineære funksjoners kritiske begreper, kommer de muligens i større grad til sin rett for denne oppgaven.

4.4 Oppgave 4 – situasjonsbeskrivelse til graf

I oppgave 4 (se appendiks nr. 4), skal elevene representere en kontekstsituasjon med hastighetsøkning i bil til graf. De 54 elevene som ble utsatt for utdypende og verifiserende tilbakemeldinger som intervensjon ($M = 0.8, SD = 0.407$), sammenlignet med de 47 elevene i kontrollgruppen ($M = 0.53, SD = 0.504$) demonstrerte signifikant forskjellig gjennomsnittskår i oppgave 4, $t(88.241) = 2.872, p = .005$. For oppgave 4 er det rimelig å støtte den alternative hypotesen. Altså er det en gjennomsnittlig prestasjonsforskjell mellom testgruppene. Med en gjennomsnittlig forskjell på 27% ligger forskjellen tett opptil resultatene fra oppgave 3. Det kan være nyttig å observere at effektstørrelsen er moderat ($g = 0.59$).

4.5 Oppgave 5 – situasjonsbeskrivelse til graf

I oppgave 5 (se appendiks nr. 5), skal elevene representere en kontekstsituasjon med vekst av en spire over tid som graf. De 54 elevene som ble utsatt for utdypende og verifiserende tilbakemeldinger som intervensjon ($M = 0.70, SD = 0.461$), sammenlignet med de 47 elevene i kontrollgruppen ($M = 0.47, SD = 0.504$) demonstrerte signifikant forskjellig gjennomsnittskår i oppgave 5, $t(94.048) = 2.437, p = .017$. Ved å observere effektstørrelsen ($g = 0.48$) ser jeg også at effekten av de utdypende tilbakemeldingene har en beskjeden effekt på den eksperimentelle gruppen.



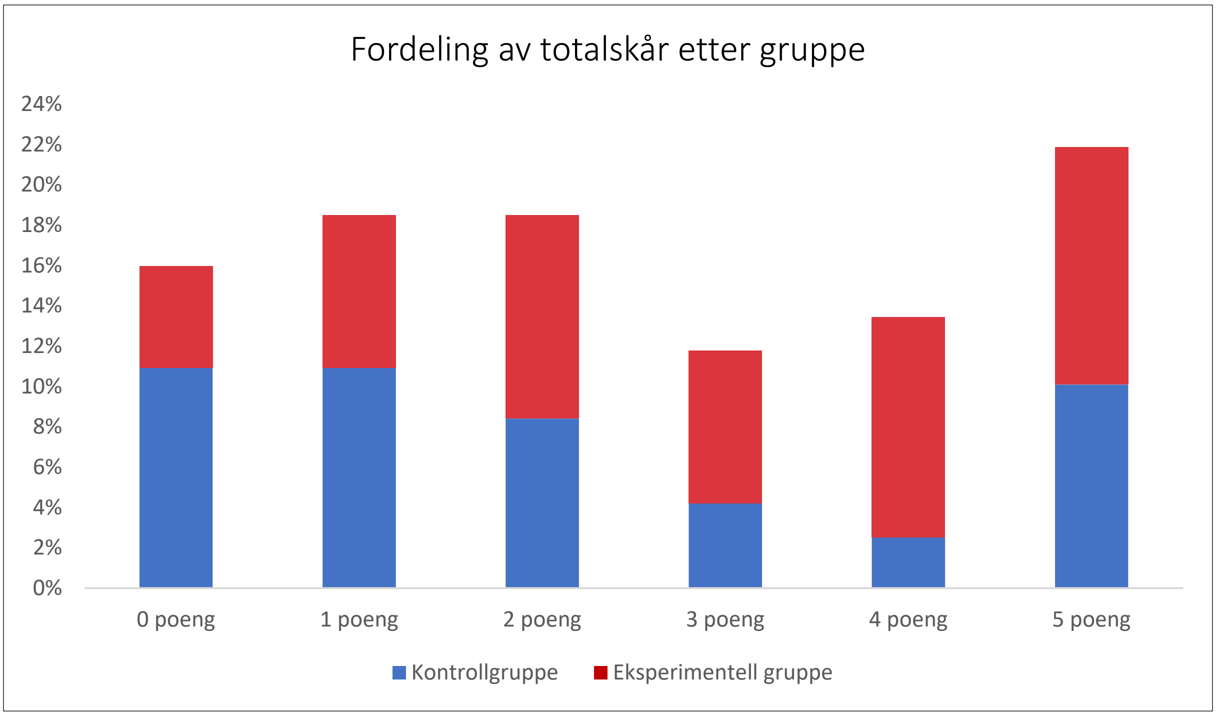
Figur 11, gjennomsnittskår på hver enkelt oppgave etter gruppe (egen produksjon).

Merk at hele 18 elevers besvarelser markeres som «missing values» i siste oppgave relativt til første, altså ni fra hver testgruppe. Gjennom testens gang forsvant disse 18 elevene underveis som følge av teknisk svikt i internettilgangen på skolene eller ved at enkeltelever valgte å hoppe over oppgaver slik at de forble ubesvart. I tillegg til disse 18 var det fire elever som valgte å hoppe over samtlige oppgaver, men disse er ikke medregnet i analysen i noen av oppgavene. Totalt gjør det at $\frac{22}{123} = 17{,}9\%$ av elevene har testbesvarelser som inneholder minst én ubesvart oppgave.

4.6 Oppgavene som helhet

Det vil i tillegg til en delvis analyse av elevenes prestasjoner på hver enkelt oppgave være nyttig å undersøke hvordan den eksperimentelle gruppen og kontrollgruppen har prestert på testen som helhet. I dette tilfellet telles en elevs totalskår opp til en maksimumsverdi på 5. Også her ble oppgaver som var besvart med en tidsbruk under 5 sekunder eller som forble ubesvart markert som manglende data. Totalt var det kun 4 elever som fikk alle sine besvarelser markert som manglende, og alle disse hadde i praksis hoppet gjennom oppgavene uten forsøk på besvarelse. Dette gjorde at selv om en bruker eksempelvis har besvart én oppgave korrekt, og fått resterende oppgaver markert som manglende, vil det ene poenget telle med i totalskåren.

Ved å anta samme hypotesetest som gjelder for de individuelle oppgavene forøvrig kan en t-test benyttes også i dette tilfellet. De totalt 63 elevene som ble utsatt for utdypende og verifiserende tilbakemeldinger som intervensjon ($M = 2.89, SD = 1.657$), sammenlignet med de totalt 56 elevene i kontrollgruppen ($M = 2.14, SD = 1.853$) demonstrerte signifikant forskjellig gjennomsnittskår i testen som helhet, $t(117) = 2.319, p = .023$. Samlet sett finner jeg derfor belegg for å støtte den alternative hypotesen om at det faktisk er en prestasjonsforskjell mellom de to testgruppene. Effektstørrelsen kan beskrives som beskjeden ($g = 0.43$), men samtidig åpenbar.



Figur 12, andelsfordeling av totalskår testgruppe (egen produksjon)

Et interessant funn når det gjelder den samlede totalskåren på tvers av gruppene fremstiller stolpediagrammet i figur 12. Diagrammet viser andelen som oppnår en gitt totalskår etter gruppen de er tildelt. Deskriptivt sett er andelen elever som får en samlet skår på 3 eller 4 poeng i den eksperimentelle gruppen henholdsvis er 2 og 3.67 ganger større enn kontrollgruppen. For den eksperimentelle gruppens del kan det altså late til at elever som vanligvis ville fått til én eller to oppgaver, nå mestret flere av dem. I tillegg er det en relativt stor gruppe elever som ser ut til å oppnå totalt 5 poeng nærmest uavhengig av hvilken gruppe de er plassert i. Utdypende tilbakemeldinger fra programvaren kan muligens være til spesiell hjelp for elever som typisk presterer på middels måloppnåelse.

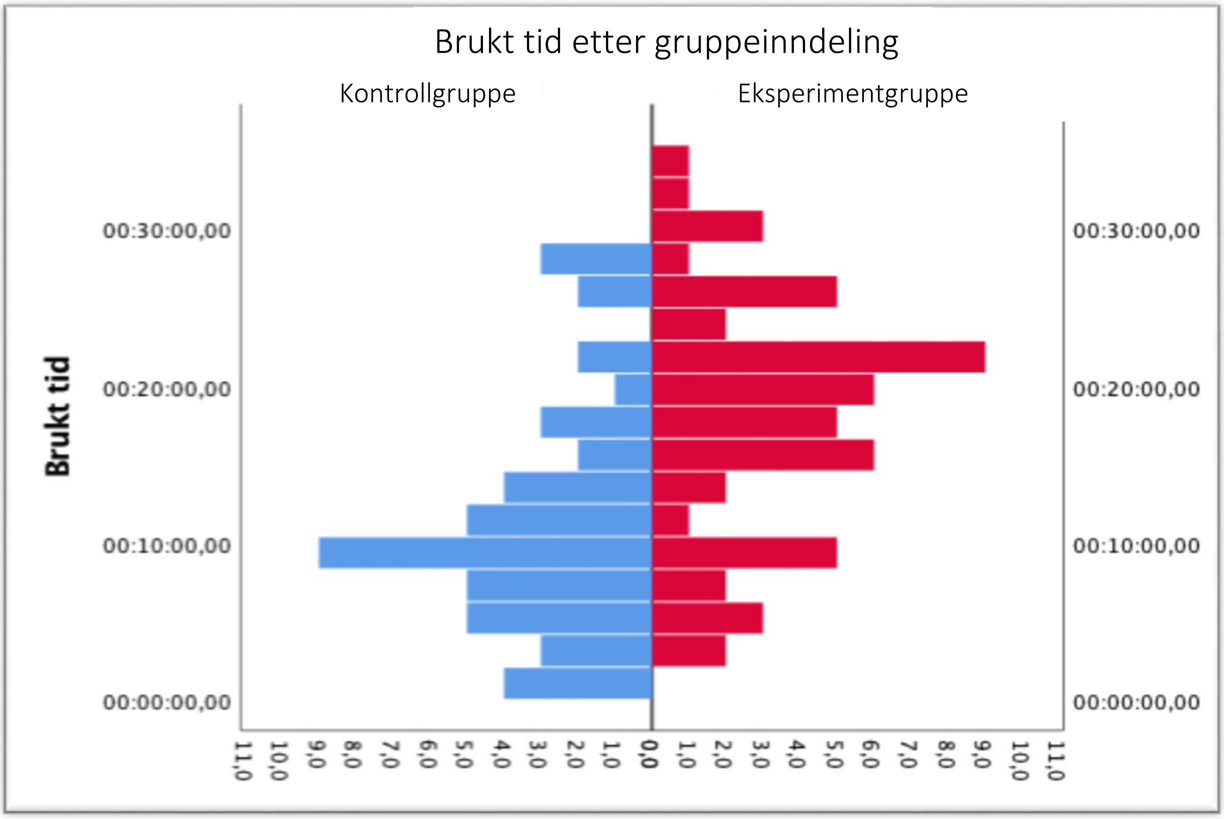
4.7 Tidsbruk i de ulike gruppene

Matistikk som testplattform gir muligheten til å presist beregne tiden hver enkelt elev bruker på oppgavene og testen som helhet. Elevenes tidsbruk kan være interessant å betrakte av ulike grunner. Tidsbruk er et aspekt som kan hjelpe til å skape noen av de gjennomsnittlige skårforskjellene, og som ikke nødvendigvis kan attribueres til tilbakemeldingenes innhold alene. Om en elev bruker mer tid gjennom testen kan det bety at de leser tilbakemeldingene eller generelt bruker mer tid på oppgavene.

For å rettferdig sammenligne elevenes tidsbruk på testen som helhet valgte jeg å utelate alle testbesvarelser som inneholdt minst én manglende oppgave. Tidligere i metodekapittelet definerte jeg manglende besvarelser som oppgaver som var hoppet over, eller av andre grunner ikke var forsøkt besvart. Målet i denne omgang er å undersøke hvor lang tid en elev bruker på alle oppgavene til sammen, og dersom en elev kun har besvart tre av fem oppgaver vil da tiden bli forholdsvis lavere enn en som har forsøkt seg på samtlige av dem. Som konsekvens er ni elever utlatt fra henholdsvis den eksperimentelle gruppen ($n = 54$) og kontrollgruppen ($n = 47$). For å undersøke om det er en forskjell i gjennomsnittstiden som er brukt på testen på tvers av testgruppene utførte jeg en t-test. Dette forutsetter igjen at det i forkant blir fremlagt en hypotesetest:

> H_0: Det er ingen statistisk signifikant forskjell i tidsbruk på testen mellom kontrollgruppen og den eksperimentelle gruppen ved $\alpha = 0.05$.
>
> H_1: Det er statistisk signifikant forskjell i tidsbruk på testen mellom kontrollgruppen og den eksperimentelle gruppen ved $\alpha = 0.05$.

De totalt 54 elevene som ble utsatt for utdypende og verifiserende tilbakemeldinger som intervensjon ($M = 00:18:10, SD = 00:07:40$), sammenlignet med de totalt 47 elevene i kontrollgruppen ($M = 00:11:35, SD = 00:07:30$) demonstrerte signifikant forskjellig tidsbruk i testen som helhet, $t(100) = 4.37, p < .001$. Samlet sett finner jeg derfor belegg for å støtte den alternative hypotesen om at det faktisk er en forskjell i tidsbruk mellom de to testgruppene. Jeg finner at den gjennomsnittlige forskjellen ligger på 6 minutter og 35 sekunder, altså et betydelig avvik. Forskjellen i tidsbruk i de to testgruppene kommer til syne i figur 13. Hva som er årsaken til denne forskjellen i tidsbruk kan muligens attribueres til to ulike faktorer.

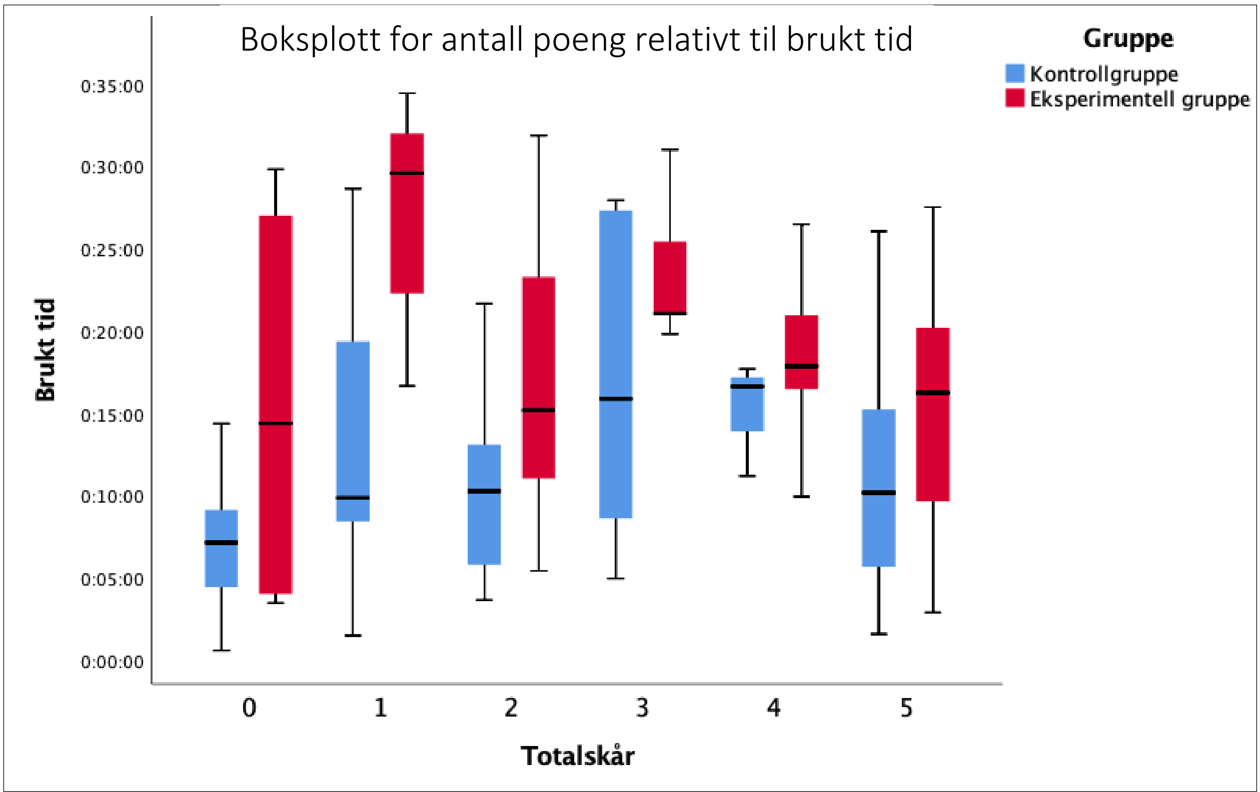


Figur 13, populasjonspyramiden viser hvordan tendensen i tidsbruk utfolder seg (egen produksjon).

Den første faktoren kan være at elevene i eksperimentgruppen bruker mer tid som følge av å måtte lese alle tilbakemeldingene. Litteraturen påpeker at en stor fare ved å gi skriftlige tilbakemeldinger er at de fort kan bli for lange, og dermed forbli uleste (Narciss & Huth, 2002; Shute, 2008). Ettersom eksperimentgruppen bruker 57% mer tid på testen virker det dog sannsynlig at de faktisk har lest tilbakemeldingene, og at mye av den ekstra tidsbruken kan attribueres til lesing.

Den andre faktoren kan beskrives som en mulighet for at de tar seg litt bedre tid til å tenke seg om i arbeid med oppgavene generelt. Å bruke litt ekstra tid til å lese oppgavens spørsmål på nytt eller undersøke hva som konkret er galt med grafen eleven har produsert, kan se ut til å øke sannsynligheten for å prestere bedre. Det er altså mulig å peke på tidsbruken og kanskje innholdet i tilbakemeldingene som viktige faktorer som kan øke den eksperimentelle gruppens prestasjoner relativt til kontrollgruppen.

Dersom jeg parer den totale tiden som er brukt mot den totalt oppnådde skåren i testen, kan jeg se tendenser mot at gruppen som ble utsatt for formative tilbakemeldinger bruker mer tid på tvers av deres prestasjoner i testen. Boksplottet i figur 14 illustrerer hvordan tidsbruken fordeler seg etter oppnådd totalskår på tvers av testgruppene.



Figur 14, boksplott som viser fordelingen av tidsbruk relativt til oppnådd totalskår i testen. Strekene i boksplottene viser medianen, og plottene er segmentert etter testgruppene (egen produksjon).

Fra et overordnet perspektiv er det ikke en klar korrelasjon mellom oppnådd totalskår og tidsbruk. Likevel finner jeg noen interessante faktorer hva angår tidsbruk for segmentene av elever fordelt etter oppnådd totalskår og testgruppe.

Blant de som får ingen poeng totalt sett er det mulig å oppdage at tidsbruken ikke overraskende er ganske lav for kontrollgruppen. For denne gruppen er tiden elevene bruker på testen totalt, homogent samlet rundt det laveste sjiktet. Elevene som oppnår null poeng er dermed de som vier minst tid til oppgavene, men også det største delsegmentet i kontrollgruppen. Det virker kanskje selvsagt, men om en ikke våger å investere nok tid i de matematiske oppgavene kan det fort resultere i slurv og generelt dårlige prestasjoner. Dersom en elev i utgangspunktet har lite motivasjon for eller interesse av å bryne seg på testen, gjenspeiler det seg altså i deres tidsbruk og dermed prestasjoner.

Til motsetning, og kanskje noe uventet, er elevene som oppnår null poeng i den eksperimentelle gruppen, svært heterogent spredd i sin tidsbruk på testen. Til tross for at de oppnår ingen poeng ser de likevel ut til å bruke mer tid enn sin motpart i kontrollgruppen. Noe av det samme mønsteret kan spores blant elevene som oppnår ett poeng i totalskår, med forskjell i at elevene i kontrollgruppen er mer spredd hva angår tidsbruk. Disse elevene ser ut til å vie mer tid til testen. Elevene som oppnår ett poeng i den eksperimentelle gruppen er overraskende nok også delsegmentet av elever som bruker mest tid på testen. Det er også for disse elevene at det er størst diskrepans i tidsbruk mellom testgruppene.

Blant elevene som oppnår to poeng i kontrollgruppen har de faktisk en noe lavere tidsbruk sammenlignet med de som oppnår ett poeng. Dette gjelder for begge testgruppene, men igjen bruker den eksperimentelle gruppen mer tid. Ser en til elevene som oppnår tre poeng kan en oppdage at tidsbruken har økt noe, og nok en gang bruker den eksperimentelle gruppen mer tid. Dersom en forstår hva oppgaven etterspør, og skjønner hvordan en oversetter en gitt representasjon av en funksjon til graf, gir det også rom for å bruke kortere tid. For kontrollgruppen sin del, er andelen som oppnår tre poeng svært lav – kun 3%, noe som kan forklare hvorfor boksplottet er så homogent samlet for dette delsegmentet.

Den eksperimentelle gruppen som oppnår fire poeng bruker også mindre tid på testen sammenlignet med de som oppnår tre poeng, men her er diskrepansen i tidsbruk på tvers av testgruppene ikke nevneverdig stor.

Dette mønsteret gjentar seg for delsegmentene av elever med høy funksjonskompetanse, som oppnår maksimal totalskår. Tidsbruken går ned sammenlignet med de som oppnår 3 poeng, men forskjellen i tidsbruk mellom testgruppene er noe større. Blant de som oppnår 4 poeng er forskjellen i tidsbruk relativt liten, men likevel marginalt høyere for eksperimentgruppen. For elevene som oppnår maksimal skår på 5 poeng, kan en se at tidsbruken er relativt lav for begge gruppene. Disse elevene mestrer emnet de testes i, noe som gjenspeiler seg i at tidsbruken for kontrollgruppen er sammenlignbar med de som oppnår 1 og 2 poeng. Igjen bruker eksperimentgruppen noe mer tid enn kontrollgruppen.

5 Drøfting

I dette kapittelet vil jeg drøfte funnene i analysekapittelet opp mot teori om oversettelse av lineære funksjoner til grafisk representasjon, effekt av tilbakemelding på elevsvar og ulike metoder for design av automatiske tilbakemeldingssystemer i matematisk læring. Kapittelet vil gi et svar på problemstillingen: «*Hvordan påvirker de automatiske formative tilbakemeldingene elevenes prestasjoner i oversettelsesoppgaver i lineære funksjoner?*». Drøftingskapittelet vil også diskutere utfordringer knyttet til design av automatiske tilbakemeldingssystemer. Både med tanke på skalering for praktisk bruk til å jobbe med matematikk, og formatet tilbakemeldingene serveres i. Tekstuelle beskjeder er ikke nødvendigvis den eneste måten å gi tilbakemeldinger digitalt på.

5.1 Effekten av automatiserte utdypende tilbakemeldinger

Her vil jeg drøfte resultatene fra oppgavene individuelt og samlet sett. I tillegg vil tidsbruken på tvers av testgruppene diskuteres.

5.1.1 Oppgave 1 og 2 - plotting

Tabell 3 viser i praksis et funn jeg fryktet som konsekvens av å fokusere på lineære funksjoners relevante begreper for grafisk konstruksjon i tilbakemeldingene. Den lokale oversettelseshandlingen plotting krever som nevnt tidligere ikke eksplisitt kjennskap til begrepene som brukes. Dette kommer til uttrykk gjennom at oppgave 1 og 2 viser henholdsvis lave og svakt negative effekter på den eksperimentelle gruppens prestasjoner relativt til kontrollgruppen, (se tabell 3 under). Forskjellene i gjennomsnittskår på disse to oppgavene er dog ikke-signifikante, og enhver forskjell mellom gruppene kan dermed sies å komme til syne av tilfeldigheter. Derfor kan jeg konkludere med at tilbakemeldingene hverken har hatt positiv eller negativ effekt på deres prestasjoner.

Oppgave	***1***	***2***
Snittskår eksperimentgruppe	$M = 0.65$	$M = 0.42$
Snittskår kontrollgruppe	$M = 0.61$	$M = 0.47$
Effektstørrelse	$g = .082$	$g = -.10$

Tabell 3, snittskår og effektstørrelse for oppgave 1 og 2

Resultatene fra oppgave 1 og 2 (se appendiks nr. 1-2), skiller seg påfallende ut i sammenligning med oppgavene som krever den globale oversettelseshandlingen skissering, altså oppgave 3-5.

Dersom jeg hadde fulgt Narciss (2004; 2013) og Narciss & Huth (2002) sine anbefalinger nøyere ved å relatere de utdypende tilbakemeldingene til feilkildene til kognitive og prosedurale ferdigheter som kreves av oppgaven, er det godt mulig en større differense mellom testgruppenes prestasjoner ville blitt mer synlig. Oppgave 1 og 2 baserer seg på oversettelseshandlingen plotting, men tilbakemeldingene gir informasjon om elementer som definerer grafens utseende. Som nevnt i metodekapittelet, lot jeg likevel testens helhetlige læringsmål – å forstå oppbyggning av en lineær graf basert på dets kritiske elementer stigningstall og konstantledd komme i første rekke. Det virker sannsynlig at å få en tilbakemeldingen "*Du har funnet rett skjæringspunkt med y-aksen, men du må finne grafens stigningstall. Hvordan kan du se på tabllen hva stigningstallet er?*", når eleven står fast på eksempelvis oppgave 1, kan virke forvirrende. For å finne stigningstallet til en lineær funksjon basert på dets tabell krever det en omregning – et unødvendig steg å gjennomføre når eleven kun trenger å flytte grafen slik at det passer med neste ordnede par fra tabellen. I tillegg vil jeg kritisere tilbakemeldingene jeg har designet til oppgave 1 og 2 for å være noe upresise. Det er ikke nødvendigvis mulig å *«[...] se på tabellen hva stigningstallet er»* når stigningstallet maskeres som endringen i y-verdi relativt til endringen i x-verdi i tabellene.

Jeg vil uansett påpeke at gjennomsnittsprestasjonene for oppgave 2 er overraskende lave, spesielt tatt i betraktning at plottingoppgaver regnes for å være mindre kompliserte enn skisseringoppgaver (Adu-Gyamfi, Bossé, & Stiff, 2012; Bossé, Adu-Gyamfi & Cheetham, 2011). At henholdsvis 42% av elevene i den eksperimentelle og 47% av elevene i kontrollgruppen får til oppgave 2 er lavt sammenlignet med oppgave 1. Faktorer som innføring av desimaltall, flere ordnede par i tabellen og at enkelte av parene inneholdt negative tall kan være forklarende her. Altså kan større mengde forvirrende fakta ha vært bidragsytende til at elevene i begge gruppene på snitt presterer dårligere på oppgave 2.

I tillegg beskriver Bossé et al. (2011) hvordan elever typisk mestrer oppgavetyper og -former som de er kjent med fra før. Hvilke representasjoner de har arbeidet mest med vil følgelig påvirke hvordan de presterer på ulike type oppgaver. Enkelte lærere i testklassene kommenterte at tabellene i oppgave 1 og 2 var orientert som rader fremfor kolonner, et aspekt de mente var uvanlig. Rent designmessig hadde det i etterpåklokskapens navn antagelig vært lettere å forholde seg til en horisontalt orientert tabell. Skjermene jeg benyttet meg av for å utforme oppgavene har større oppløsning og størrelse enn Chromebookene elever i Trondheim bruker. Den lave oppløsningen førte til at elevene måtte rulle skjermen sin opp og ned for å betrakte tabellen og deretter grafikkvinduet de representerte grafen i. Dersom en ser mot de tre øvrige oppgavene er det betydelig større forskjeller mellom gruppene.

5.1.2 Oppgave 3, 4 og 5 - skissering

Fra analysen og tabell 4 under blir det tydelig at oppgavene som krever oversettelseshandlingen skissering (3, 4 og 5 – se appendiks nr. 3-5) viser tydelige, signifikante prestasjonsforskjeller på tvers av testgruppene. I arbeid med å skissere en graf fra algebraisk eller rent tekstuell informasjon, kreves det at elevene kan arbeide med begreper som stigningstall og skjæringspunkt med y-aksen for å representere grafen (Bossé et al., 2011). I slike tilfeller kommer fokuset på begrepene kanskje mer til sin rett. At resultatet er så tydelig i favør til eksperimentgruppen for disse tre oppgavene, understreker poenget til Narciss (2004; 2013) og Narciss & Huth (2002) om at tilbakemeldingene bør adressere de konseptuelle kunnskapene og de prosedurale ferdighetene de krever.

Oppgave	***3***	***4***	***5***
Snittskår eksperimentgruppe	$M = 0.61$	$M = 0.8$	$M = 0.7$
Snittskår kontrollgruppe	$M = 0.31$	$M = 0.53$	$M = 0.47$
Effektstørrelse	$g = .62$	$g = .59$	$g = .48$

Tabell 4, snittskår og effektstørrelse for oppgave 3, 4 og 5

Det er også mulig å argumentere for at elevene har blitt mer kjent med det digitale koordinatsystemet, måten tilbakemeldingene leveres på, eller generelt lært seg å representere en graf ut ifra oppgaven gjennom oppgavenes gang. Disse faktorene kan være med på å påvirke skårforskjellen mellom testgruppene for oppgave 3, 4 og 5. Muligheten for at de kognitive kravene senkes for mye som følge av for førende tilbakemeldinger for disse oppgavene, er også en faktor som kan bidra.

Generelt ser det ikke ut til at oppgavenes tiltenkte vanskelighetsgrad har vært slik jeg i utgangspunktet antok. Jeg designet testen ut ifra at første oppgave vil være den minst kompliserte, hvorpå en økning i kompleksitet i oppgavene ville resulterer i gradvis dårligere prestasjoner frem mot oppgave fem. For eksperimentgruppen er de antatt vanskeligste oppgavene (tekstoppgave 4 & 5) oppgavene de presterer best på. Resultatet står dermed i kontrast med Adu-Gyamfi et al. (2012) og Bossé et al. (2011) sine påstander om at globale oversettelseshandlinger som skissering vil være mest førende for oppgavenes antatte vanskelighetsgrad.

5.1.3 Total prestasjonseffekt i testen

Cohen et al. (2018) går i bresjen for å akkompagnere eller til og med bytte ut tester som viser statistisk signifikans med informasjon om effektstørrelser. Cohen et al. (2018) sine tolkninger av effektstørrelser ved hjelp av tommelfingerregler hevder at svak effekt $= 0 - 0.20$, $beskjeden$ effekt $= 0.21 - 0.50$, $moderat$ effekt $= 0.51 - 1$ og $sterk$ effekt > 1 for Cohens d og Hedges g. Samtidig peker Ellis (2019) til at disse koeffisientene kan tolkes som svak effekt $= 0.20$, moderat effekt $= 0.50$ og sterk effekt $= 0.80$. I metodekapittelet tok jeg opp hvordan beregning og bruk av effektstørrelser på mange måter er irrelevante uten øvrig sammenligning med andre studier (Cohen J. , 1988; Ellis, 2019; Hutchison & Styles, 2010). De øvrige nevnte generelle tolkningene bør settes i sammenheng for å forstås relativt til andre studier på maskingitte tilbakemeldingers effekt.

Tilbakemeldinger blir generelt ansett som et positivt tilskudd til læring, men Kluger & DeNisi (1996) hevder i sin metaundersøkelse at den gjennomsnittlige effektstørrelsen av skriftlige tilbakemeldinger er beskjeden ($d = 0.41$). De konstaterer at funnene av effektstørrelsene er svært heterogene, mye på grunn av stor variasjon i testutvalgenes størrelser, men også fordi de observerte store forskjeller i hvordan tilbakemeldingene ble servert, og hvordan effekten ble målt. I enkelte studier måles prestasjoner, som i denne undersøkelsen, mens andre ser på effekten tilbakemeldinger har på læring av ulike tema over tid. Samtidig påpeker Shute (2008) at omlag en tredjedel av alle studier på tilbakemeldinger viser til *negative* effekter.

Tabell 5 viser at den effektstørrelsen de utdypende tilbakemeldingene har på eksperimentgruppens totalskår i testen er beskjeden til moderat ($g = 0.43$), men likevel tilnærmet lik det andre studier finner (Hattie, 2009; 2011; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Van der Kleij, Feskens & Eggen, 2015).

Oppgave	1	2	3	4	5	Totalskår
Effektstørrelse *(g)*	.082	-.10	.62	.59	.48	.43
Sig.	$p = .626$	$p = .626$	$p = .002$	$p = .005$	$p = .017$	$p = .023$

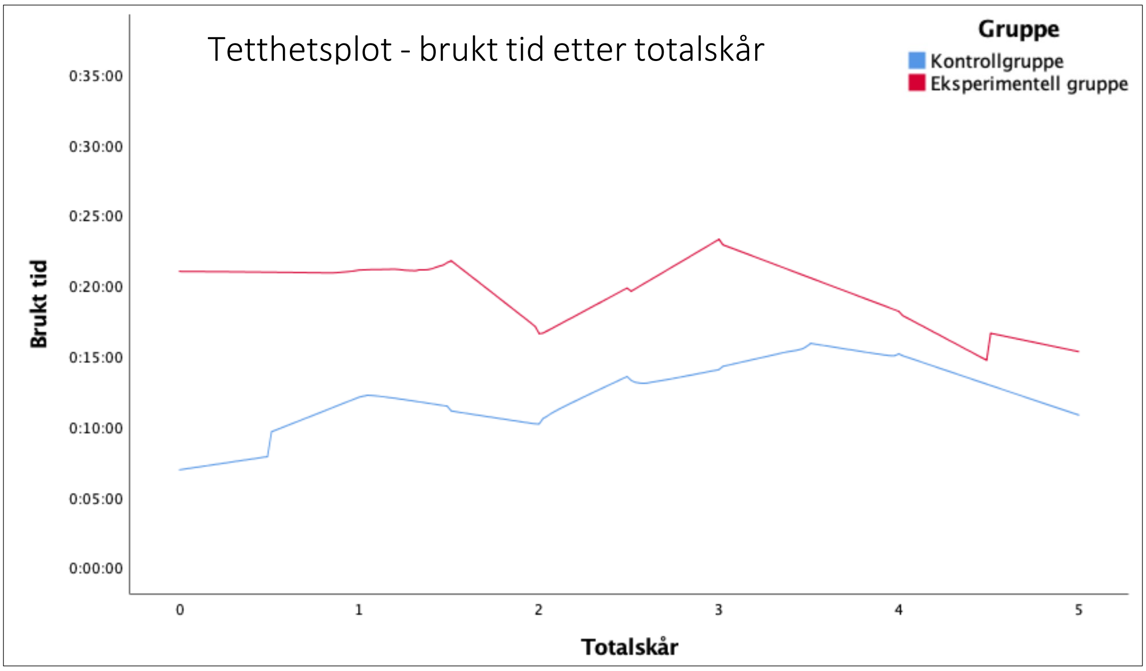
Tabell 5, effektstørrelser og signifikansnivå for resultatene på hver enkelt oppgave, og testen som helhet (egen produksjon).

Van der Klei, Feskens & Eggen (2015) viser i sin metaundersøkelse på effekten av tilbakemeldinger gitt av en datamaskin på læring og prestasjoner, at utdypende tilbakemeldinger fremstår som mest effektivt ($g = 0.49$). De finner også at rent verifiserende tilbakemeldinger ikke ser ut til å ha noen særlig effekt ($g = 0.05$). Dersom programvaren gir det korrekte svaret som tilbakemelding, slik Kikora og Khan Academy gjør, er effekten større enn verifisering, men mindre enn utdypning ($g = 0.32$). De bemerker at effekten av datamaskingitte tilbakemeldinger ser ut til å være større for matematikk sammenlignet med samfunnsfagene, naturfagene og språkfagene.

Resultatene fra totalskår i denne undersøkelsen er også sammenlignbart med bruk av Intelligent Turoring Systems (ITS), slik som Gyldendal Maximum Smartøving. Skoler som bruker ITS som verktøy for å arbeide med lekser, eller i timer orientert rundt bruken av en slik plattform oppnår totalt en effektstørrelse på $d = 0.48$ i studier på deres effekt (Hattie, 2011). Forskerne målte hvilken effekt bruken av ITS hadde på flere klasser relativt til kontrollgrupper på den samme og ulike skoler. Resultatene de måler kommer fra en standardisert test på 9. trinnselever, der resultatene ble sammenlignet for å produsere en effekt på det de kaller læring. Slik sett føyer undertegnedes undersøkelse inn i materien av funn som allerede eksisterer (Kluger & DeNisi, 1996; Van der Kleij et al., 2015). Som motsetning har denne undersøkelsen kun sett på rene prestasjoner i en og samme test.

5.1.4 Effekt på elevenes tidsbruk

Det var forventet at tidsbruken til eksperimentgruppen ville være noe høyere enn for kontrollgruppen, slik tetthetsplottet i figur 15 viser. At tidsforskjellen kom til å være så stor som den viste seg å være fremstår likevel som noe overraskende.



Figur 15, tetthetsplot med epanechnikov kernel for å glatte ut grafene. Figuren viser sammenhengen mellom brukt tid på testen og oppnådd totalskår – rød linje representerer den eksperimentelle gruppen, og den blå linjen representer kontrollgruppen (egen produksjon)

I analysen presenterte jeg to mulige faktorer som kunne forklare forskjellen i tidsbruk – at elevene bruker tid på å lese tilbakemeldingene, og at elevene bruker mer tid på å arbeide med og revurdere sine svar underveis, altså en form for utholdenhet og motivasjon. Den første hypotesen fremstår som sannsynlig, men samtidig ikke selvsagt, tatt i betraktning Shute (2008) sine advarsler om at for lange og komplekse tilbakemeldinger ofte ikke blir lest av elever. Figur 16 på neste side viser en potensiell fare som designet av programvaren presenterer; nemlig at tidligere gitte tilbakemeldinger stables over hverandre. En slik mengde tekst og informasjon kan virke krevende å forholde seg til, og dermed risikere å ikke bli lest.

Det er også godt mulig at elevene bruker mer tid på å tenke seg om for hver oppgave – at de utviser større utholdenhet og muligens motivasjon for arbeid med oppgavene. Spesielt for elevene som oppnår 0 eller 1 poeng i testen er tidsbruken for eksperimentgruppen betraktelig større enn kontrollgruppen. At tidsbruk-forskjellen er størst for disse elevene står i kontrast til Narciss (2004) sine funn. Hun beskriver at elever som allerede har høy motivasjon for faget og gode prestasjoner fra før, også er elevene som ser ut til å ha best utbytte av feilrelaterte tilbakemeldinger fra en datamaskin hva gjelder motivasjon og prestasjon.

Figur 16, utsnitt fra tilbakemeldingsskjermen elevene ser når de svarer på oppgaven de arbeider med (egen produksjon).

Om jeg antar at stor diskrepans i tidsbruk i favør til eksperimentgruppen for lavt presterende elever kan komme delvis av økt utholdenhet og motivasjon, vil dette være en styrke med maskingitte tilbakemeldinger generelt. På den annen side kan den store forskjellen i tidsbruk for lavt presterende elever også influeres av sen lesehastighet.

Tidsbruk i testen og dets sammenheng med motivasjon og utholdenhet for å jobbe med oppgavene er uansett vanskelig å fastslå med presisjon for denne undersøkelsen. Det viser også at sammenhengen mellom motivasjon, utholdenhet og ulike typer maskingitte tilbakemeldinger bør undersøkes nærmere (Hattie, 2009; Narciss, 2004).

5.1.5 For mye eller for lite hjelp?

Det kan reises spørsmål om hvorvidt de utdypende tilbakemeldingene senket oppgavenes kognitive krav, slik Tangen (2018) hevder at blant annet Kikoras tilbakemeldingsmetode gjør. Kikora gir som nevnt i kapitel 2.6.2 tilbakemeldinger i form av verifisering og løsning på oppgaven eleven jobber med, så det er ikke overraskende at denne påstanden kan diskuteres for Kikora som nettplattform. De utdypende tilbakemeldingene i denne undersøkelsen (se appendiks 7.6) gir tvert imot ikke løsningen på oppgavene. De gir informasjon om *hva* eleven har gjort feil, hint til mulige strategier eleven kan bruke og metakognitive motspørsmål. Faktoren som muligens kan bidra mest til å senke det kognitive nivået for enkelte oppgaver vil være hintene til mulige strategier. Jeg vil likevel påstå at disse hintene ikke er for ledende, men ofte noe vage. Til tross for dette er det likevel en sannsynlighet for at noen av de utdypende tilbakemeldingene kan ha vært for førende. Dette ville ikke vært noe problem om jeg hadde testet tilbakemeldingenes påvirkning på læring over tid, men i en prestasjonstest vil spørsmålet om senkning i kognitive krav være relevante.

5.2 Kritisk blikk på studiens generaliserbarhet

Er resultatene som analysene i denne undersøkelsen viser til generaliserbare for en øvrig populasjon utover testdeltakerne? Spørsmålet vil i stor grad være avhengig av hvorvidt elevene i de to testgruppene faktisk er sammenlignbare i utgangspunktet. Ettersom jeg ikke gjennomførte en pre-test på elevene er det heller ikke mulig å påstå at forskjellene i testskår faktisk dukker opp som følge av intervensjonen den eksperimentelle gruppen ble utsatt for. Kanskje elevene i eksperimentgruppen i utgangspunktet ville oppnådd bedre skår i utgangspunktet? En av måtene jeg forsøkte å demme opp for denne trusselen mot resultatets validitet var å dele opp hver klasse i en eksperiment -og kontrollgruppe. På denne måten er hver enkelt deltagende klasse en del av begge testgruppene.

Om jeg ser til fordelingen av totalskår fordelt etter de fire skolene og de to testgruppene kan vi se i tabell 6 under at det er store forskjeller mellom dem. For alle utenom skole 3 kan det skimtes en markant skårforskjell i favør eksperimentgruppen på tvers av testgruppene i totalskår. Det kan påstås at elevene som havnet i eksperimentgruppen i hver klasse i utgangspunktet ville oppnådd større totalskår, men da de ble tilfeldig fordelt innad i klassene, virker dette usannsynlig. Skolene oppnår varierende resultater, men felles for de fleste av dem påvirker de utdypende tilbakemeldingene resultatet i positiv retning.

Skole	*Skole 1*	*Skole 2*	*Skole 3*	*Skole 4*
Eksperimentgruppe	$M = 2.87$	$M = 4.11$	$M = 2.72$	$M = 2.25$
Kontrollgruppe	$M = 1.52$	$M = 3.2$	$M = 2.76$	$M = 1.67$

Tabell 6, snittskår fordelt på de ulike skolene

5.3 Andre implikasjoner av automatiske tilbakemeldinger og digitale læringssystemer.

Bruken av en datamaskin til å automatisk bedømme og vi tilbakemeldinger på elevsvar fører med seg noen endringer i hvordan en arbeider for å oppnå læring. Digitale systemer er ikke statiske slik for eksempel en lærebok i matematikk er, og det gir noen muligheter, men setter også noen begrensninger for praktisk bruk.

5.3.1 Skalering av algoritmiske tilbakemeldingssystemer

En av de tydeligste utfordringene med å utvikle og bruke et system slik jeg har skissert gjennom denne undersøkelsen, utmaler Rønning (2017) når han påpeker hva som er Maple T.A. sin største utfordring. Å produsere formative tilbakemeldinger gjennom systemet er i utgangspunktet basert på at et menneske manuelt må forutsi elevers potensielle feilkilder på forhånd. Det gjør at generering av oppgaver blir tidkrevende og dyrt arbeid. En kan altså stille seg spørsmålet om effekten et slikt system har på elevenes læring kan rettferdiggjøres relativt til utfordringene som er iboende dets design. Å skalere opp et slikt system til å brukes på et stort antall elever for deres øvings skyld, kan derfor bli dyrt og tidkrevende.

Denne undersøkelsen produserte en test med kun fem oppgaver innenfor et begrenset domene av oversettelse fra ulike funksjonsrepresentasjoner til graf. Oppgavene i seg selv er slik sett ganske begrenset i kompleksitet. De krever ingen spesiell kunnskap i av elevene utover å kunne generere en graf til en gitt situasjon. Når effektstørrelsen på totalskår kan sies å være beskjeden ($g = 0.43$), viser det at automatiske tilbakemeldingssystemer basert på algoritmisk oppdagelse av feil med feilrelatert veiledning har sin plass, men også har et stort forbedringspotensial.

5.3.2 Oppgavenes karakter

En utfordring som blir synliggjort gjennom å undersøke ulike private aktørers tjenester, enten det er Khan Academy, Kikora, Maple T.A. eller Gyldendals Smartøving, er oppgavene de leverer med mulighet for automatisk tilbakemelding. Narciss (2004) peker på at de er gjerne begrensede i natur – ofte i form av «multiple choice»-oppgaver eller ved at elevene skal beregne et algebraisk uttrykk, eller svar i numerisk form. Oppgavene i denne undersøkelsen baserer seg på en annen metodikk, nemlig at elever skal tilpasse en graf. Likevel fremstår også de som noe enkle og prosedyreorienterte individuelt sett (Valenta, 2016). Om en betrakter alle de fem oppgavene samlet, kan det dog sies at de sammen med tilbakemeldingene har et høyere kognitivt krav.

Uansett blir den grunnleggende teknikken for å evaluere elevenes svar også oppgavenes akilleshæl. Den teknologiske utfordringen det er for en datamaskin å kunne evaluere mer sammensatte problemstillinger som krever komplekse svar med rom for tolkning, gjør at oppgavene ofte blir begrensede og prosedyreorienterte av karakter (Egelandsdal et al., 2019; Morlandstø et al., 2019; Narciss, 2004; Valenta, 2016).

Samtidig åpner bruk av digitale plattformer opp for mulighet til repetisjon og øving ved at en stor mengde oppgaver med tilfeldig genererte variabler kan løses av elever og studenter (Barana & Marchisio, 2016). Tilbakemeldingene kan deretter tilpasses disse genererte variablene og dermed gi tilbakemelding på oppgaven, gitt at den ikke overstiger en viss kompleksitet. Som et eksempel, kan en rekke oppgaver, ikke ulikt oppgave 3 i denne undersøkelsen, genereres med ulike verdier for uttrykkets konstantledd og stigningstall.

5.3.3 Multimodalitet i tilbakemeldingene

En fare forbundet å presentere tilbakemeldinger eksklusivt som tekst er at det kan lede til å kognitivt overbelaste elevene med for mye informasjon. Dette aspektet kan endres ved bruk av digitale verktøys multimodalitet (Narciss & Huth, 2002; Sacristán, et al., 2010). Tilbakemeldinger trenger ikke låses til å være rent tekstuelle elementer, men kan presenteres som tale, med lyder, filmklipp og bevegelige bilder eller ikoner (Ostrow & Heffernan, 2014; Shute, 2008). Enkelte private aktørers plattformer som Khan Academy og Gyldendals Maximum Smartøving utnytter spesielt muligheten til å gi eksempler, definisjoner og løsninger gjennom korte filmklipp (Gyldendal Akademisk, 2019; Murphy et al., 2014). Khan Academy fikk sitt gjennombrudd ved Salman Khans korte filmsnutter der han går gjennom utallige eksempler og oppgaver.

En kan da lure på i hvilken grad bruken av filmer og andre multimedieformater som tilbakemeldinger eventuelt påvirker elevene. I denne undersøkelsen valgte jeg å benytte tekststrenger som tilbakemelding, men det utelukker ikke at det finnes mer effektive formater. Ostrow & Heffernan (2014) hevder i sitt eksperiment der 138 åttendeklassinger i arbeid med å lære det Pytagoreiske teorem, presterer bedre $(M = 0.77, SD = 0.43)$ når de får tilbakemeldinger på oppgavene i form av videoklipp, fremfor rent tekstuelle tilbakemeldinger $(M = 0.63, SD = 0.50)$. Resultatet deres er *ikke* signifikant $(p = .143)$, og effektstørrelsen er beskjeden $(g = 0.32)$. Artikkelen peker likevel på at elevene selv foretrekker tilbakemeldinger i filmatisk format fremfor gitt som tekststrenger av maskinen. En overveldende andel på 83% av elevene i undersøkelsen deres forteller at de ønsker tilbakemeldinger gitt som korte filmklipp. Ostrow & Heffernan (2014) hevder også at elevenes motivasjon for å arbeide med oppgavene økte signifikant med inkluderingen av videotilbakemeldinger. Det kan altså virke som læringsutbytte kanskje kan øke ved å inkludere videoklipp som tilbakemelding, men det er hos elevenes motivasjon en finner størst gevinst ved deres bruk (Ostrow & Heffernan, 2014).

Det er altså mulig at denne undersøkelsen har gått glipp av en marginal økning i prestasjoner ved å ikke tilby tilbakemeldinger med videoklipp.

5.3.4 Bruksområdet til automatiske tilbakemeldingssystemer

I hvilken situasjon kan automatiske tilbakemeldingssystemer brukes? Denne undersøkelsen har brukt en testsituasjon for å måle konsekvensen det har på prestasjoner, men automatiske tilbakemeldingssystemer og digitale læringsplattformer har langt større potensiale (Narciss, 2013). Hattie (2009) skriver at læreren er den primære agenten for å gi tilbakemeldinger i en klasseromsituasjon, men at datamaskiners vurderinger også har sin plass i forbindelse med læring. Læreren blir ikke gjort overflødig, men en godt utformet digital læringsplattform med tilbakemeldinger gir derimot nok et verktøy læreren kan ha i sitt arsenal.

De fleste digitale læringsplattformer brukes som øvingsoppgaver til lekser, ekstraarbeid eller «flipped clasroom»-aktiviteter (omvendt undervisning) (Morlandstø et al., 2019; Murphy et al., 2014; Rønning, 2017). Spesielt i hjemmeaktiviteter vil det oppstå et fravær av en lærer som kan tilby tilbakemelding på elevarbeid. Dette gjør at en datamaskins evaluering av oppgaver får en naturlig nisje som syntetisk mentor. Enkelte elever har foreldre som kan hjelpe dem med matematikkens utfordringer til langt opp i videregående skolenivå, mens andre ikke er så heldige. På denne måten kan innføring av automatiske tilbakemeldingssystemer kanskje være med på å jevne ut forskjellene som skapes av foreldrenes ulike utdanningsnivå.

I tillegg har digitale læringsplattformer generelt en stor styrke når det gjelder prognostisk evaluering (Barana & Marchisio, 2016). Systemer som Gyldendals Maximum Smartøving kan avsløre hull i elevers forståelse i forkant av undervisning, slik at læreren lettere kan forberede timer. Samtidig kan enkelte adaptive læringsalgoritmer som Smartøving og deres tilbakemeldingsmetode føre til et svært individuelt orientert fokus (Egelandsdal et al., 2019). For at tilbakemeldingene og algoritmen skal virke etter sin hensikt er den avhengig av at elever arbeider med den alene. Når de arbeider individuelt vil programvaren forsøke å tilpasse oppgavene relativt til det antatte nivået til eleven (Egelandsdal et al., 2019; Morlandstø et al., 2019). En slik tilbakemeldingsmetode fasiliterer dermed ikke til samarbeid, noe som igjen kan peke mot at automatisert tilbakemelding og matematiske e-læringsplattformer ofte blir brukt til terping, øving og som hjemmelekser.

5.4 Konklusjon

Metoden jeg har brukt for å designe tilbakemeldingssystemet viser at feilrelatert veiledning kan brukes for å oppnå prestasjonsforskjell i testen. Dette svarer på delspørsmålet om «Hvordan kan det designes automatiske formative tilbakemeldinger til oppgaver om oversettelse fra tabellform, algebraisk- og situasjonsrepresentasjon til grafisk representasjon av lineære funksjoner?»

Funnene fra analysen og drøftingen viser at de utdypende tilbakemeldingene totalt sett har en positiv effekt på eksperimentgruppens prestasjoner, men samtidig at tilbakemeldingene bør reflektere oversettelseshandlingen i oppgavene. Jeg finner ingen signifikant prestasjonsforskjell i de to oppgavene som krever den lokale oversettelseshandlingen plotting. Til motsetning finner jeg moderat prestasjonsforskjell i de tre oppgavene som krever den globale oversettelseshandlingen skissering – kanskje på grunn av tilbakemeldingenes fokus på begreper som styrer de aktuelle grafenes utseende. Fokuset på begreper kan ha blitt for stort relativt til oversettelseshandlingene for oppgave 1 og 2.

Generelt er effekten på prestasjoner av de datamaskingitte utdypende tilbakemeldingene sammenlignbar med det annen lignende forskning finner. I tillegg ser jeg gjennom elevenes tidsbruk på testen at de faktisk leser tilbakemeldingene. Slik sett besvares altså det andre delspørsmålet «Hvordan påvirker de automatiske formative tilbakemeldingene elevenes prestasjoner i oversettelsesoppgaver i lineære funksjoner?».

I tillegg ser det ikke ut til å være en klar sammenheng mellom oversettelseshandlingenes antatte kompleksitet og elevenes prestasjoner. Oppgavene som krever oversetting fra tabell til grafisk representasjon, spesielt oppgave to, viste seg å være overraskende vanskelig for elevene.

5.4.1 Videre forskning

Det finnes mange veier å gå med tanke på den videre forskningen på automatiske vurderingssystemer og e-læringsverktøy generelt for matematikk. Tatt i betraktning at denne undersøkelsen hovedsakelig har hentet inspirasjon fra en studie som i 2020 er 18 år gammel, og at det i mellomtiden har dukket opp mange private bedrifter med nye typer digitale læringsverktøy, bør dette temaet forskes mer på (Narciss & Huth, 2002). Spesielt longitudinelle studier for flere alderssegmenter enn 10. klasseelever bør inkluderes.

Jeg ser for meg en studie der systematisk bruk av enten et automatisk tilbakemeldingssystem utviklet i forskningsøyemed, eller en av de private aktørenes programmer, med pre-test og en standardisert prøve i et gitt tema ved slutten av et semester kan gi mer utfyllende informasjon. Denne undersøkelsen har kun sett på prestasjoner i en snever og svært kort test. Selv om effekten av de utdypende tilbakemeldingene på prestasjoner kan sies å være ikke-neglisjerbare, viser denne undersøkelsen kun dets effekt på prestasjoner. Det sier ikke så mye om hvordan et automatisk vurderingssystem med formative tilbakemeldinger påvirker elevers læring over tid.

Det har også skjedd en stor utvikling i programvareutvikling, datamaskiners regnekraft og smarte algoritmers kapasitet. Om en går ut ifra antagelsen om at denne progresjonen ikke stopper med det første, ser jeg for meg at ulike digitale læringsplattformer vil ha en økende innflytelse på skolene verden over. Elementer av lærerarbeidet kan assisteres ved hjelp av datamaskiners unike evner til å katalogisere informasjon og analysere mønster i dem. ITS som Gyldendals Smartøving gir antageligvis bare en smakebit på hvilke informasjonssystemer som kan komme til å prege skolehverdagen i de kommende årene.

De digitale plattformene har i tillegg en nåværende kundedekning som med stor fart har spredd seg til mange brukere. Khan Academy brukes allerede av 2,4 millioner elever i det amerikanske skolesystemet, men dette og andre lignende verktøy vil antageligvis bli bedre og mer presise over tid (Murphy et al., 2014). Jeg vil derfor hevde at det å forstå hvordan automatiske vurderingssystemer, formative som rent summative, og e-læringsverktøy generelt påvirker våre elever er svært kritisk.

6 Litteraturliste

Adu-Gyamfi, K., & Bossé, M. J. (2014). Processes and reasoning in representations of linear functions. *International Journal of Science and Mathematics Eduaction, 12*, ss. 167-192.

Adu-Gyamfi, K., Bossé, M. J., & Stiff, L. V. (2012). Lost in Translation: Examining Translation Errors Associated With Mathematical Representations. *School Science and Mathematics Vol. 112, No. 3*, ss. 152-170.

Ainley, J. (2001). Adjusting to the newcomer - Roles for the computer in mathematics classrooms. I P. Gates, *Issues in Mathematics Teaching* (ss. 166-181). London, UK: Routledge.

Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences 11(1)*, ss. 25-62.

Barana, A., & Marchisio, M. (2016, juni). Ten good reasons to adopt an automated formative assessment model for learning and teaching Mathematics and scientific disciplines. *Procedia - Social and Behavioral Sciences, No. 228*, ss. 608-613.

Barana, A., Marchisio, M., & Rabellino, S. (2015). Automated assessment in mathematics. I S. I. Ahamed, C. K. Chang, W. Chu, I. Crnkovic, P.-A. Hsuing, G. Huang, & J. Yang, *The 39th Annual IEEE Computer Software and Applications Conference (COMPSAC)* (ss. 670-671). Los Alamitos, California, US: IEEE Computer Society.

Blackman, R. T. (2014). Using Maple, Maple TA, Mathlynx and the tablet PC to teach undergraduate calculus. *Electronic Proceedings of the Twenty-Sixth Annual Conference on Technology in Collegiate Mathematics*, ss. 10-16.

Blum, W., & Ferri, R. B. (2009). Mathematical Modelling: Can It Be Taught And Learnt? *Journal of Mathematical Modelling and Application, Vol. 1, No. 1*, ss. 45-58.

Bossé, M. J., Adu-Gyamfi, K., & Cheetham, M. R. (2011, Oktober). Assessing the Difficulty of Mathematical Translations: Synthesizing the Literature and Novel Findings. *International Electronic Journal of Mathematics Education – IΣJMΣ, Vol.6, No.3*, ss. 113-133.

Brito, I., Figueiredo, J. M., Flores, M. A., Jesus, A., Machado, A., Malheiro, G. J., . . . Vaz, E. (2009). Using e-learning to self regulate the learning process of mathematics for engineering students. I C. A. Bulucea, V. Mladenov, E. Pop, M. Leba, & N. Mastorakis, *Recent Advances in Applied Mathematics. Proceedings of the 14th International*

Conference on Applied Mathematics (MATH-09) (ss. 165-169). Puerto de la Cruz, Spania: WSEAS.

Bueie, H. (2011). *Geogebra for lærere.* Oslo, Norge: Universitetsforlaget.

Char, B. (2011, August 1). Developing questions for Maple TA using Maple library modules and non-mathematical computation. *Drexel University, Department of Computer Science*, ss. 1-49.

Chiu, M. M., Kessel, C., Moschkovic, J., & Muñoz-Nuñez, A. (2001, juni 7). Learning to Graph Linear Functions: A Case Study of Conceptual Change. *Cognition and Instruction, No. 19(2)*, ss. 215-252.

Clement, J., Lockhead, J., & Monk, G. S. (1981). Translation difficulties in learning mathematics. *American Mathematical Monthly, No. 88*, ss. 287-290.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, US: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education, 8. utgave.* New York, US: Routledge Taylor & Francis Group.

Creswell, D. J. (2018). *Research Design - Qualitative, Quantitative, and Mixed Methods Approaches.* New York, US: Sage Publications inc.

Dahl, H. A. (2014, juni). Digital læringsressurs -et bidrag til a styrke matematikkopplæring? - Masteroppgave i IKT-støttet læring. Oslo, Norge. Hentet fra https://oda-hioa.archive.knowledgearc.net/bitstream/handle/10642/2144/dahl_hilde_aske_.pdf?sequence=2&isAllowed=y

Doorman, L. M., & Gravemeijer, K. P. (2009). Emergent modeling: discrete graphs to support the understanding of change and velocity. *ZDM Mathematics Eduaction, Vol. 41*, ss. 199-211.

Drijvers, P. (2012). Digital Technology in Mathematics Education: Why It Works (Or Doesn't). *Proceedings of the 12th International Congress on Mathematics Education*.

Duval, R. (2002). The cognitive analysis of problems of comprehension in the learning of mathematics. *Mediterranean Journal for Research in Mathematics Education, 1(2)*, ss. 1-16.

Duval, R. (2006). The cognitive analysis of problems of comprehension in the learning of mathematics. *Educational Studies in Mathematics, 61*, ss. 103-131.

Egelandsdal, K., Hansen, C. S., Ness, I. J., & Wasson, B. (2019). *Adaptiv læring i matematikk. Empirisk rapport om Multi Smart Øving i grunnskolen.* Bergen, Norge: SLATE Research Report.

Ellis, P. D. (2019). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results.* Cambridge, UK: Cambridge University Press.

Even, R. (1989). Factors Involved in Linking Representations of Functions. *Journal of Mathematical Behaviour, No. 17(1)*, ss. 105-121.

Even, R. (1990). Subject Matter Knowledge for Teaching and the Case of Functions. *Educational Studies in Mathematics, Vol. 21, No. 6*, ss. 521-544.

Flaa, P. D. (2009). Evaluering av Kikora: Pedagogiske implikasjoner for Kikora som digitalt læringsverktøy i matematikk. Oslo: Lærelyst A/S.

Greenhow, M. (2015, juni). Effective computer-aided assessment of mathematics; principles, practice and results. *Teaching Mathematics and Its Applications*.

Guba, E. G. (1981). ERIC/ECTJ Annual Review Paper: Criteria for Assessing the Trustworthiness of Naturalistic Inquiries. *Educational Communication and Technology, Vol. 29, No. 2*, ss. 75-91.

Gyldendal Akademisk. (2019). Maximum Smartøving. Oslo, Norge.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* London, UK: Routledge.

Hattie, J. (2011). *Visible Learning for Teachers.* London, UK: Taylor & Francis Ltd.

Hattie, J., & Timperley, H. (2007, Mars 1). The Power of Feedback. *Review of Educational Research*, ss. 81-112.

Hedges, L. V. (1981, juni). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics, Vol. 6, No. 2*, ss. 107-128.

Hewitt, D. (2012). Young students learning formal algebraic notation and solving linear equations: are commonly experienced difficulties avoidable? *Educational Studies in Mathematics 81:*, ss. 139-159.

Hutchison, D., & Styles, B. (2010). *A Guide to Running Randomised Controlled Trials for Educational Researchers.* United Kingdom, The Mere, Upton Park, Slough, Berkshire SL1 2DQ: National Foundation for Educational Research (NFER).

Kaput, J. J. (1992). Technology and mathematics education. I I. D. Grouws, *Handbook of Research on Mathematics Teaching and Learning* (ss. 515-556). Information Age Publishing.

Klegseth, M. R. (2018). An assessment of non-standardized tests of mathematical competence for Norwegian secondary school using Rasch analysis . ss. 1-94.

Kluger, A. N., & DeNisi, A. S. (1996, Mars). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin, Vol. 119, No. 2*, ss. 254-284.

Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning, Vol. 22*, ss. 24-33.

Kramarski, B., & Zeichner, O. (2001). Using Technology to Enhance Mathematical Reasoning: Effects of Feedback and Self-Regulation Learning. *Education Media International*, ss. 77-82.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in Written Instruction: The Place of Response Certitude. *Educational Psychology Review, 1(4)*, ss. 279-308.

Landers, R. N. (2015). Psychological Theory and the Gamification of Learning. I T. Reiners, & L. C. Wood, *Gamification in Education and Business* (ss. 165-186). Curtin, Australia: Springer.

Leinhardt, G., Zaslavsky, O., & Stein, M. (1990). Functions, Graphs, and Graphing: Tasks, Learning, and Teaching. *Review of Educational Research, Vol. 60, No. 1*, ss. 1-64.

Lysø, K. O. (2017). *Sannsynlighetsregning og statistisk metodelære.* Bergen: Caspar Forlag.

Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology*, ss. 449-471.

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition Learning*, ss. 85-97.

Morlandstø, N. I., Hansen, C. S., Wasson, B., & Bull, S. (2019). *Aktivitetsdata for vurdering og tilpasning: Sluttrapport.* Bergen, Norge: Centre for the Science of Learning & Technology (SLATE).

Mory, E. H. (1996). Feedback in research. I D. H. Jonassen, *Handbook of research for educational communications and technology* (ss. 916-956). New York, US: Simon & Schuster Maxmillan.

Murphy, R., Gallagher, L., Krumm, A., Mislevy, J., & Hafter, A. (2014). *Research on the Use of Khan Academy in Schools.* California, US: SRI - Education. Finansiert av The Bill & Melinda Gates Foundation og MiT, Boston.

Narciss, S. (2004, Februar). The Impact of Informative Tutoring Feedback and Self- Efficacy on Motivation and Achievement in Concept Learning. *Experimental Psychology*, ss. 1-25.

Narciss, S. (2013, Juni). Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review, No. 23*, ss. 7-26.

Narciss, S., & Huth, K. (2002). How to design informative tutoring feedback for multi-media learning. *Psychology of Learning and Instruction*, ss. 1-16.

Nitsch, R., Fredebohm, A., Bruder, R., Kelava, A., Naccarella, D., Leuders, T., & Wirtz, M. (2015). Students' Competencies in Working with Functions in Secondary Mathematics Education - Empirical Examination of a Competence Structure Model. *International Journal of Science and Mathematics Eduaction 13*, ss. 657-682.

Nortvedt, G. A., & Pettersen, A. (2016). Kapittel 6 - Matematikk. I M. Kjærnsli, & F. Jensen, *Stø kurs - Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015* (s. 229). Oslo: Universitetsforlaget.

O'Callaghan, B. R. (1998, januar). Computer-Intensive Algebra and Students' Conceptual Knowledge of Functions. *Journal for Research in Mathematics Education, Vol. 29, No. 1*, ss. 21-40.

Ostrow, K. S., & Heffernan, N. T. (2014). Testing the Multimedia Principle in the Real World: A Comparison of Video vs. Text Feedback in Authentic Middle School Math Assignments. *Proceedings of the 7th International Conference on Educational Data Mining*, ss. 296-299.

Pontoppidan, M., Keilow, M., Dietrichson, J., Solheim, O. J., Opheim, V., Gustafson, S., & Andersen, S. C. (2018). Randomised controlled trials in Scandinavian educational research. *Educational Research, Vol. 60, no. 3*, ss. 311-335.

Rønning, F. (2017, Februar). Influence of computer-aided assessment on ways of working with mathematics. *Teaching Mathematics and Its Applications, Vol. 36*, ss. 94-107.

Robson, C. (2002). *Real World Research - A Resource for Social Scientists and Practitioner-Researchers. 2. utgave.* Massachusets, US: Blackwell Publishing.

Sacristán, A. I., Calder, N., Rojano, T., Santos-Trigo, M., Friedlander, A., Meissner, H., . . . Perrusquía, E. (2010). The Influence and Shaping of Digital Technologies on the Learning – and Learning Trajectories – of Mathematical Concepts. I C. Hoyles, & J. B. Lagrange, *Mathematics Education and Technology - Rethinking the Terrain* (ss. 179-226). NY, US: Springer Science + Business Media.

Sfard, A. (1991). On the Dual Nature of Mathematical Conceptions: Reflections on Processes and Objects as Different sides of the Same Coin. *Educational Studies in Mathematics 22*, ss. 1-36.

Sfard, A. (2007, Desember 5). When the Rules of Discourse Change, but Nobody Tells You: Making Sense of Mathematics Learning From a Commognitive Standpoint. *Journal of the Learning Sciences Vol. 16, No. 4*, ss. 565-613.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research, Vol. 78, No. 1*, ss. 153-189.

Sierpinska, A. (1992). On Understanding the Notion of Function. I E. Dubinsky, & G. Harel, *The Concept of the Function: Elements of Pedagogy and Epistemology* (ss. 25-58). US: Notes and Reports series of The Mathematical Association of America .

Skemp, R. R. (1971). *The Psychology of Learning Mathematics.* Harmondsworth, England: Penguin Books.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teacher 77*, ss. 20-26.

Tangen, P. H. (2018). *Digitale oppgavers kognitive krav.* Trondheim, No: NTNU.

Usiskin, Z. (2012, juli). What does it mean to understand some mathematics? *12th International Congress on Mathematical Education*, ss. 1-20.

Utdanningsdirektoratet. (2014). *Læreplan i matematikk fellesfag (MAT1-04).* Hentet fra Kompetansemål etter 10. årssteget: https://www.udir.no/kl06/MAT1-04/Hele/Kompetansemaal/kompetansemal-etter-10.-arssteget

Utdanningsdirektoratet. (2020). *Fagfornyelsen.* Hentet fra Matematikk, kompetansemål og vurdering 10. trinn: https://www.udir.no/lk20/mat01-05/kompetansemaal-og-vurdering/kv14

Valenta, A. (2016). *Kognitive krav i matematikkoppgaver.* Trondheim, Norge: Matematikksenteret - Nasjonalt senter for matematikk i opplæringen.

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015, desember). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research, Vol. 85, No. 4*, ss. 475-511.

Wiliam, D. (2011). *Embedded formative assessment.* Bloomington, Indiana, USA: Solution Tree Press.

Winkler, S., Körner, A., & Urbonaite, V. (2012). Maple T.A. in Engineering Educations. *IFAC Proceedings Volumes, Vol. 45*, ss. 906-911.

Ødegaard, J. (2016, juni). "Smart Øving vet ka æ kan" - En kvalitativ studie av hvordan elever bruker den adaptive læringsressursen Smart Øving i matematikkfaget. Masteroppgave i spesialpedagogikk. Trondheim, Norge. Hentet fra NTNU, Open: https://ntnuopen.ntnu.no/ntnu-

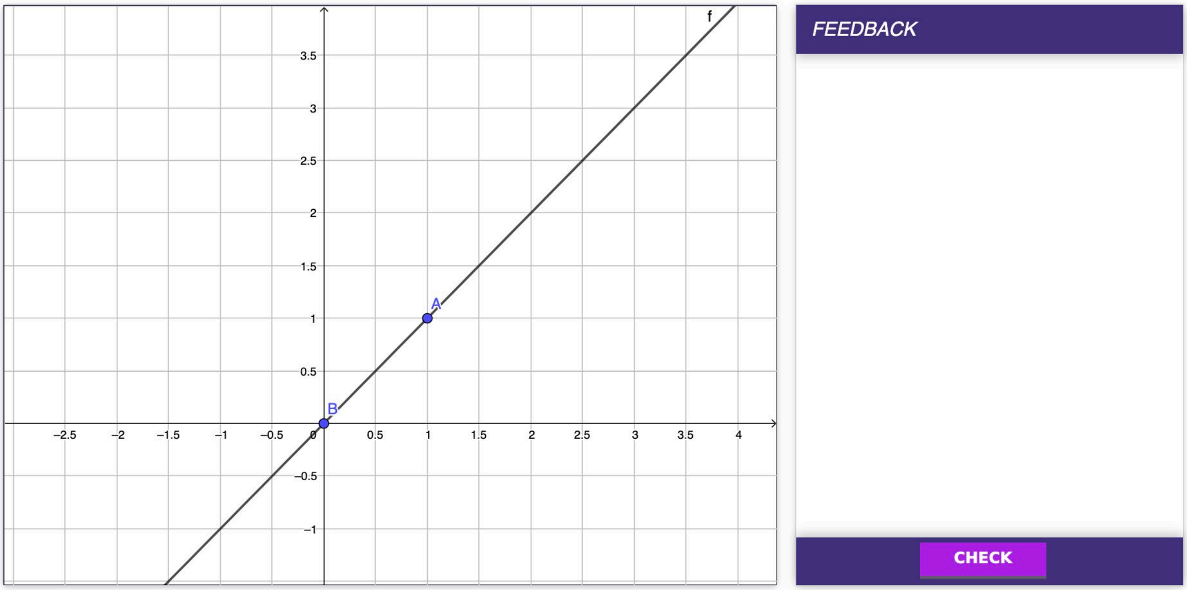
xmlui/bitstream/handle/11250/2479083/Julie%20%C3%98degaard.pdf?sequence=1&i
sAllowed=y

7 Appendiks

7.1 Oppgave 1

«Her skal du forsøke å dra i punktene A og B slik at grafen som dannes mellom dem passer med *tabellen* du ser under. Du kan holde «shift» inne og dra i aksene for å forminske eller forstørre dem, samt zoome med musepekeren eller fingrene dine om du har touch-skjerm.»

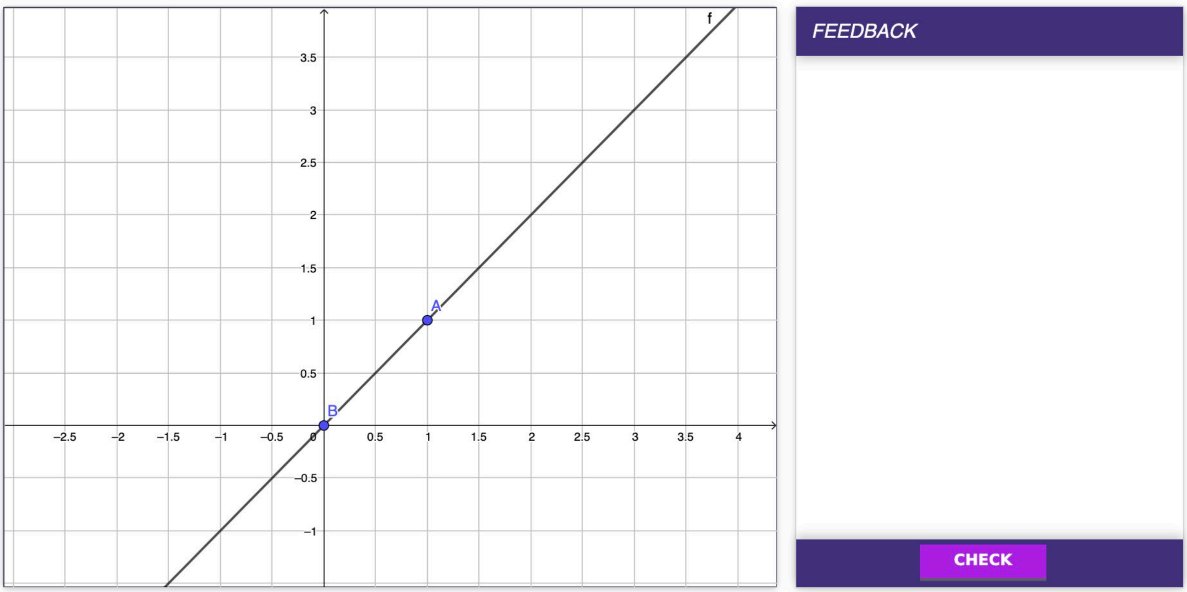
x	y
0	5
1	0
2	-5
3	-10
4	-15
5	-20



7.2 Oppgave 2

«Her skal du forsøke å dra i punktene A og B slik at grafen som dannes mellom dem passer med *tabellen* du ser under. Du kan holde «shift» inne og dra i aksene for å forminske eller forstørre dem, samt zoome med musepekeren eller fingrene dine om du har touch-skjerm.»

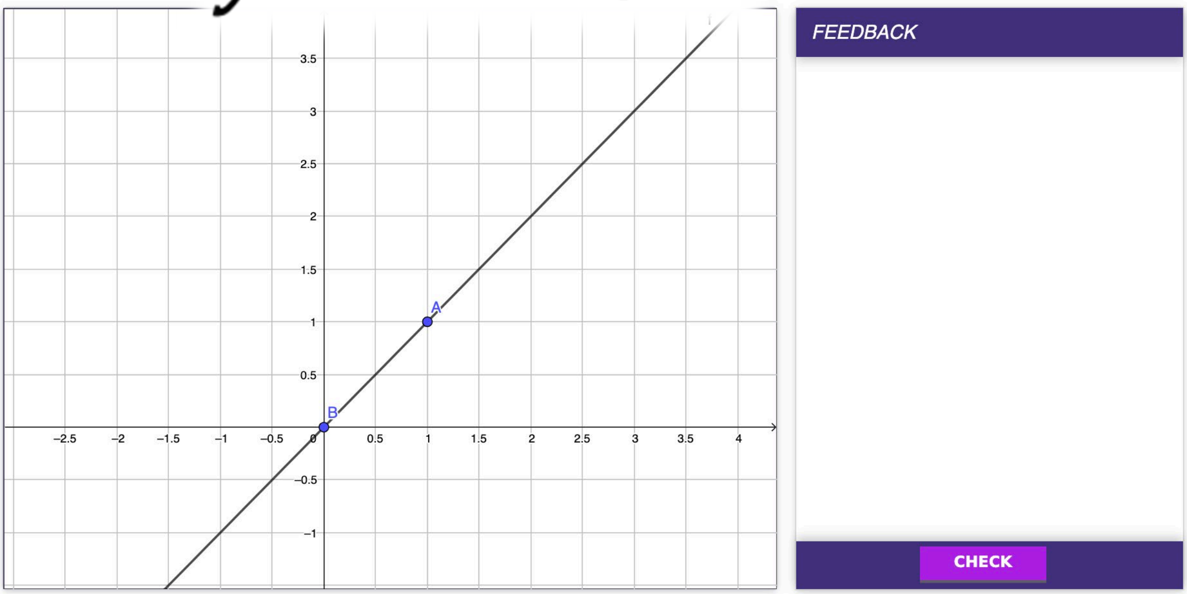
x	y
-6	-27.5
-4	-19.5
-2	-11.5
0	-3.5
2	4.5
4	12.5
6	20.5



7.3 Oppgave 3

«Her skal du forsøke å dra i punktene A og B slik at grafen som dannes mellom dem passer med *funksjonen* du ser under. Du kan holde «shift» inne og dra i aksene for å forminske eller forstørre dem, samt zoome med musepekeren eller fingrene dine om du har touch-skjerm.»

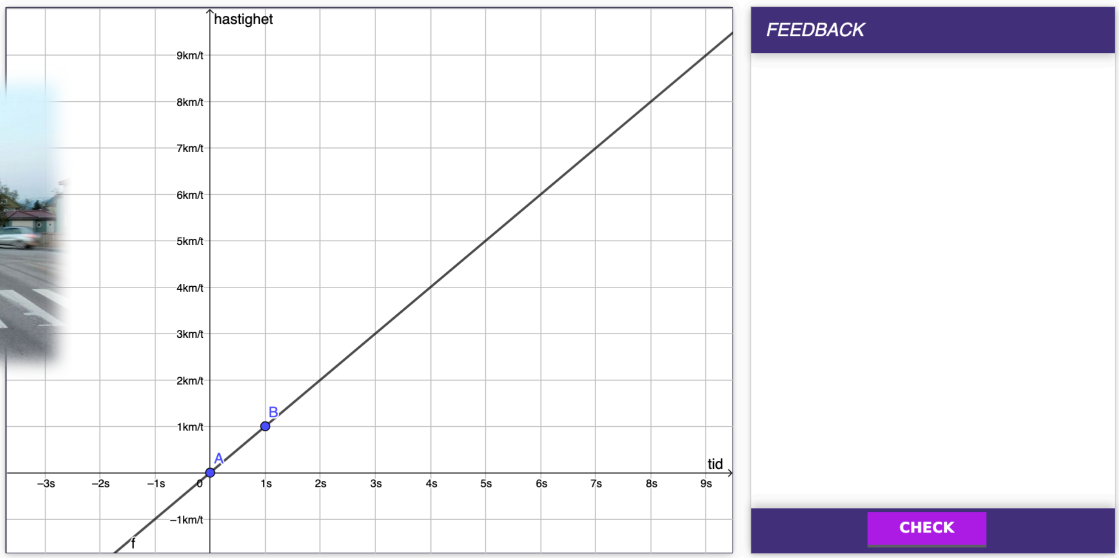
$$y = -3x + 6$$



7.4 Oppgave 4

«Du sitter i en bil ved et lyskryss. I det trafikklysene skifter til grønt setter du foten på gasspedalen.

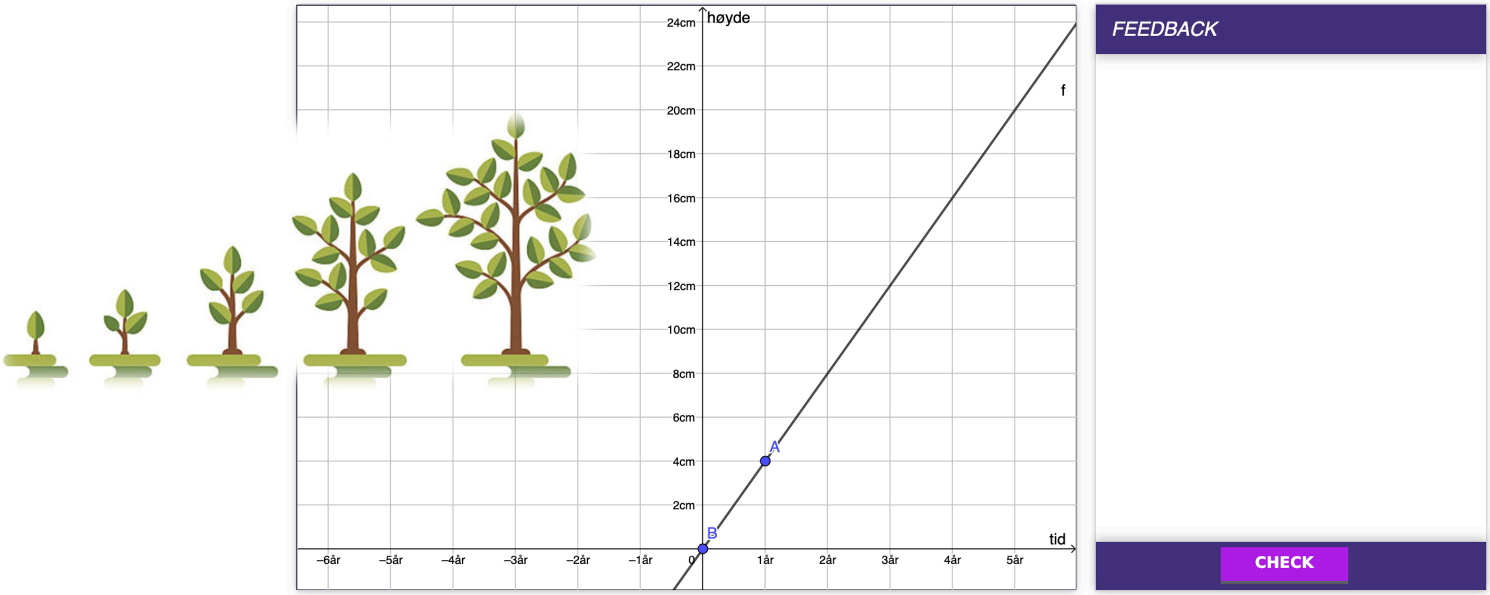
- Bilen øker sin hastighet med 10 km/t hvert andre sekund du sitter i bilen.
- Dra i punktene A og B slik at grafen som tegnes mellom dem viser bilens hastighet over tid, fra trafikklyset blir grønt.»



7.5 Oppgave 5

«Tenk deg at du vil ha en plante på eiendommen din.

- Tidlig på våren planter du et lite tre på 24 centimeter i hagen.
- Hvert år etter treet ble plantet registrerer du at det har vokst 14 centimeter.
- Tilpass grafen med punktene A og B slik at den viser hvor høyt treet ditt er etter et gitt antall år.»



7.6 Alle gitte tilbakemeldinger i programvaren og deres triggere

Trigger for tilbakemelding	Her fremkommer alle tilbakemeldinger som blir gitt, samt hvilke feiltilstander som utløser hver enkelt type feedback. Korrekt svar vil alltid defineres av en lineær funksjon $y = m + E$ når den har verdier lik en annen funksjon $y_1 = m_1 + F$ Feil som er nærmere å være korrekt vil være markert med en gul farge, mens feil som er mer alvorlige er markert med en rød farge. Disse fargene presenteres med tilbakemeldingen til elevene, og det ble på forhånd presisert at disse fargene sier dem noe om hvor nærme de er å klare oppgaven.
Oppgave 1 (tabell lik $y = -5x + 5$)	"☆Korrekt svar!☆" **når** $\boldsymbol{m_1 = m}$ **&** $\boldsymbol{E = F}$
$\boldsymbol{m = m_1}$ **&** $\boldsymbol{E\ != F}$ →	- "Du har funnet rett stigningstall, men det virker som grafen din ikke skjærer y-aksen på korrekt plass. Hvordan kan du finne skjæringspunktet i tabellen?" - "Grafens skjæringspunkt med y-aksen finner du når $x = 0$. Du bør finne skjæringspunktet, for du har allerede funnet rett stigningstall!"
$\boldsymbol{m < 0}$ **&** $\boldsymbol{E\ != F}$ →	- "Du har konstruert en graf med negativ stigning, noe som stemmer. Likevel er den enten for slak eller for bratt. Hvor mye endrer grafen seg for hver x- verdi?", - "Hvor er grafens skjæringspunkt med y-aksen? Tabellen viser hvor dette skjæringspunktet befinner seg!", - "Både skjæringspunkt med y-aksen og de ulike koordinatene som grafen beveger seg gjennom finner du i tabllen!"
$\boldsymbol{m\ != m_1}$ **&** $\boldsymbol{E = F}$ →	- "Du har funnet rett skjæringspunkt med y-aksen, men du må finne grafens stigningstall. Hvordan kan

	du se på tabllen hva stigningstallet er?", - "Du mangler bare å korrekt finne grafens stigningstall. Stigningstallet sier noe om hvor bratt grafen synker eller stiger. Vil grafen synke eller stige her?", - "Skjæringspunktet med y-aksen har du korrekt funnet til $(0, 5)$, men gjennom hvilke andre punkter bør linja bevege seg? Disse punktene finner du i tabellen over."
$\boldsymbol{m = 0} \rightarrow$	- "Grafen din har ingen stigning. Kan du se ut ifra tabellen hva stigningstallet kan være?"
$\boldsymbol{m > 0\ \&\ E\ != F} \rightarrow$	- "Grafen din stiger positivt, noe som ikke stemmer. Kan du se på tabellen hvor negativ endringen i y bør være?", - "Du mangler også å finne grafens skjæringspunkt med y-aksen. Når treffer grafen y-aksen?
Oppgave 2 (tabell lik $y = 4x - 3{,}5$)	"☆Korrekt svar!☆" **når** $\boldsymbol{m_1 = m\ \&\ E = F}$
$\boldsymbol{m = m_1\ \&\ E\ != F} \rightarrow$	- "Du har funnet rett stigningstall, men det virker som grafen din ikke skjærer y-aksen på korekt plass. Hvordan kan du finne skjæringspunktet i tabellen?", - "Grafens skjæringspunkt med y-aksen finner du når $x = 0$. Kan du finne skjæringspunktet, for du har allerede funnet rett stigning!"
$\boldsymbol{m < 0\ \&\ E\ != F} \rightarrow$	- "Grafen din har negativ stigning, noe som ikke stemmer med tabellen. Øker eller synker y-verdiene nedover i tabellen?", - "Grafen din skjærer ikke y-aksen på rett sted. Hvor i tabellen kan du se dette skjæringspunktet?", - "Både skjæringspunkt med y-aksen og de ulike koordinatene som

	grafen beveger seg gjennom finner du i tabllen!" - "Grafen din synker, noe ikke stemmer ut ifra tabellen. Kan du se hvor mye grafens y- verdi synker per x- verdi?", - "Dersom grafens stigningstall hadde vært negativt ville også y-verdiene i tabellen minket i verdi nedover."
$m \mathrel{!}= m_1$ & $E = F$ →	- "Du har funnet rett skjæringspunkt med y-aksen, men du må finne grafens stigningstall. Hvordan kan du se på tabllen hva stigningstallet er?", - "Du mangler bare å korrekt finne grafens stigningstall. Husk at stigningstallet sier noe om hvor bratt grafen synker eller stiger.", - "Igjen har du funnet rett skjæringspunkt med y-aksen, så nå mangler bare grafens stigningstall. Ser du noen likheter eller ulikheter med tabellen i denne oppgaven sammenlignet med den forrige? Bør grafen synke eller stige?"
$m = 0$ →	- "Grafen din har ingen stigning. Kan du se ut ifra tabellen hva stigningstallet kan være? Hvor mye endrer funksjonens verdi seg mellom hver x?"
$m > 0$ & $E \mathrel{!}= F$ →	- "Grafen din har positivt stigningstall, som er riktig. Kan du se på tabellen hvor mye grafen stiger i y- verdi for hver x-verdi?", - "Hvor er grafens skjæringspunkt med y-aksen? Tabellen viser hvor dette skjæringspunktet befinner seg!", - "Både skjæringspunkt med y-aksen og de ulike koordinatene som grafen beveger seg gjennom finner du i tabellen!"

Oppgave 3 (likning lik $y = -3x + 6$) $\boldsymbol{m = m_1 \,\&\, E \,! = F} \rightarrow$ $\boldsymbol{y = 2x + 6}$ **(tolkningsfeil der uttrykket tolkes som koordinater)** $\rightarrow$ $\boldsymbol{m < 0 \,\&\, E \,! = F} \rightarrow$ $\boldsymbol{m \,! = m_1 \,\&\, E = F} \rightarrow$	"☆Korrekt svar!☆" **når** $\boldsymbol{m_1 = m \,\&\, E = F}$ - "Du har funnet rett stigningstall, men det virker som grafen din ikke skjærer y-aksen på korekt plass. Hvordan vil et uttrykk som skjærer y-aksen i f.eks. y = 0 se ut?", - "Grafen din har rett stigning, men dens skjæringspunkt med y- aksen finner du i lineære funksjoner som b i $y = ax + b$. " - "Grafen skal skjære y- aksen når $y = 6$, så akkurat det er riktig. Husk at $-3x$ viser hvor mye grafen stiger/synker for hver x- verdi." - "Grafen din har negativ stigning, noe som stemmer. Kan du se på uttrykket hvor bratt den negative stigningen bør være?", - "Du har laget en graf som ikke har helt rett stigning eller skjæringspunkt med y-aksen. Ved hvilken y-verdi skjærer grafen y-aksen?", - "Stigningstallet finner du foran x og skjæringspunktet med y- aksen er leddet helt til høyre i uttrykket." - "Du har funnet rett skjæringspunkt med y-aksen, men du må finne grafens stigningstall. Hvordan kan du se på uttrykket hva stigningstallet er?", - "Ettersom grafen din skjærer y-aksen i $(0, 6)$ er dette riktig, men du mangler å finne rett stigning på grafen. Stigningstallet står i uttrykket som -3.", - "Når stigningstallet er -3 betyr det at funksjonen synker med 3 y-

$\boldsymbol{m = 0} \rightarrow$ $\boldsymbol{m > 0 \ \& \ E \ != F} \rightarrow$ $\boldsymbol{E < 0} \rightarrow$	verdier for hver x-verdi." - "Grafen din har ingen stigning, altså har den samme y-verdi for alle x-verdier og fremstår derfor som flat. Funksjonens uttrykk viser hva stigningstallet kan være.", - "Når en graf har null stigning vil den være $y = 0x + b$ - altså er det tallet foran x som bestemmer hvor fort den stiger eller synker." - "Grafen din stiger positivt, noe som ikke stemmer og skjæringspunktet med y-aksen ditt er galt.", - "Hvordan kan du se på uttrykket hva stigningstallet og skjæringspunktet med y-aksen bør være?", - "Stigningstallet finner du foran x og skjæringspunktet med y- aksen er leddet helt til høyre i uttrykket." - "Skjæringspunktet med y-aksen er negativt. Se på uttrykket $y = -3x + 6$. Kanskje du kan prøve å regne ut y blir hvis $x = 0$ for å finne det?"
Oppgave 4 (situasjonsbeskrivelse lik $y = \frac{10x}{2}$) $\boldsymbol{m = m_1 \ \& \ E \ != F} \rightarrow$ **Syntaktisk strategi - tolkning av tall fremfor tekst ($\boldsymbol{m = 10 \ / \ A = (1, 10) \ OR \ F = (10, 1)}$)** $\rightarrow$	"☆Korrekt svar!☆" **når** $\boldsymbol{m_1 = m \ \& \ E = F}$ - "Du har funnet bilens rette fart over tid, men det virker som bilen din beveget seg fremover du begynte å gasse. Hvordan kan du finne bilens startfart ved lyskrysset?", - "Dersom du står stille ved et lyskryss i det trafikklyset blir grønt og deretter akselererer vil startfarten din være $= 0$. Hvordan ser dette ut som graf?" - "Her kan du ha vært litt rask. Tar det bare 1 sekund før bilen har en hastighet på 10km/t?"

$m < 0 \rightarrow$	- "Grafen din har negativ stigning, noe som vil bety at bilen du styrer vil rygge. Kan du si ut ifra situasjonen om endringen i *farten* er negativ eller positiv?"
$m = 0 \;\&\; E = F \rightarrow$	- "Uten fartsendring blir bilen din stående helt stille ved lyskrysset. Hvor mye stiger farten med for hvert sekund?"
$m \,!= \, m_1 \;\&\; E = F \rightarrow$	- "Du har funnet rett skjæringspunkt med y-aksen, altså bilens startfart, men kan du finne hvor fort bilen endrer fart?", - "Stigningstallet til grafen blir bilens fartsøkning. Les oppgaveteksten en gang til og tenk: 'hvor fort kjører du etter 2 $sekunder$?'", - "Startfarten til bilen er fortsatt rett, men du må finne dens fart over tid. Se på aksene i grafikkfeltet. x-aksen er markert som antall sekunder som er gått, mens y-aksen viser hastigheten på bilen."
$m = 0 \rightarrow$	- "Grafen din har ingen stigning. Kan du se ut ifra oppgaven hva farten til bilen kan være? Uten stigning vil ikke farten din endre seg!"
$m > 0 \rightarrow$	- "Hvor fort kjører du? Farten du kjører i øker over tid, så hvor fort kjører du etter ett sekund?", - "Bilen du styrer øker farten, noe du har vist her. Likevel viser grafen din en hastighet som ikke helt stemmer med situasjonen. Tenk hvor fort du kjører etter 2 sekunder. Hva med 3 eller 4 sekunder?"

Oppgave 5 (situasjonsbeskrivelse lik $y = 14x + 24$)	"☆Korrekt svar!☆" **når** $\boldsymbol{m_1 = m \,\&\, E = F}$
$\boldsymbol{m = m_1 \,\&\, E \; != F} \rightarrow$	- "Du har funnet det lille treets korrekte årlige vekst, men hvordan kan du vise spirens starthøyde ved planting på grafen?", - "Ettersom du har klart å vise hvor høyt treet er over tid, mangler du bare å finne hvor høyt det var ved planting. Hvor høyt er det ved tid $= 0$?"
$\boldsymbol{m < 0} \rightarrow$	- "Grafen din har negativ stigning. Vil treet bli mindre eller større for hvert år som går?", - "Med negativ stigning vil treet bli mindre for hvert år som går."
$\boldsymbol{m = 0 \,\&\, E = F} \rightarrow$	- "Grafen din viser at det lille treet var 24 cm da det ble plantet, men samtidig at det ikke vokser. Stemmer det med beskrivelsen i oppgaven?", - "Du har funnet rett skjæringspunkt med høydeaksen. Samtidig er stigningen til grafen din flat som vil si at treet har samme høyde over tid.", - "Treet skal vokse over tid, noe du kan vise ved å få grafen til å stige. Hvor mye vil det vokse med ifølge oppgaven?"
$\boldsymbol{E = 0} \rightarrow$	- "Den forrige oppgaven tok utgangspunkt i en bil som starter stillestående. Hvor bør grafen skjære y-aksen da du ikke bare plantet et frø, men en liten plante?"
$\boldsymbol{m = 0} \rightarrow$	- "Grafen din har ingen stigning. Det vil si at plantens høyde ikke endrer seg mellom hvert år. Hvordan kan du vise det lille treets vekst med grafen? Er det noen likheter med

$m > 0 \;\&\; E \,!= F \rightarrow$ $m \,!= m_1 \;\&\; E = F \rightarrow$ **(Tolkningsfeil – bytter mellom hva som er tid (x-akse) og høyde (y-akse)**	forrige oppgave?" - "Grafen din stiger positivt, noe som gir mening med tanke på det lille treets vekst. Kan du se på situasjonen hvor raskt den bør stige?" - "Grafen din viser at planten ble plantet når det var 24 cm høyt, men hva med dets vekst? Forrige oppgave beskrev også en form for vekst - ser du noen likheter mellom denne og forrige oppgave?" - "Det kan hende du tenker litt omvendt. Grafen din krysser y-aksen på $14\ cm$, og x- aksen etter $24\ år$, hvilket betyr at det lille treet ble plantet $14\ cm$ høyt og ble gradvis mindre før det forsvant etter 24 år."

Sindre Mathias Strømnes Nordvoll

Automatiske tilbakemeldingssystemer i matematisk læring

NTNU
Norwegian University of
Science and Technology