Misoppfatninger innen betinget sannsynlighet for elever i den videregående skole

Erik Holst

Erfaringsbasert master i undervisning med fordypning i matematikk

Matematisk institutt
Universitetet i Bergen
juni 2016

Forord
Med denne oppgaven fullfører jeg et langt løp som startet 2008. Jeg vil takke veileder Runar Ile ved matematisk institutt for stor tålmodighet under prosessen.

Abstrakt
Masteroppgaven handler om ulike misoppfatninger elever på videregående skole har innen emnet betinget sannsynlighet. I teorien belyses misoppfatninger, feilslutninger og heuristikker med hovedvekt på betinget sannsynlighet. I oppgaven har jeg brukt mixed method, en blanding mellom kvantitativ og kvalitativ metode. Til den kvantitative delen har jeg brukt oppgaver innen betinget sannsynlighet utarbeidet av Diaz og Fuente. De kaller denne samlingen av oppgaver for CPR (Conditional Probability Reasoning). Jeg bruker de samme oppgavene i den kvalitative delen av masteroppgaven i oppgavebasert intervju. Jeg sammenligner mine resultat med resultat fra CPR, og studerer graden av ulike misoppfatninger, relatert til grupper som har hatt eller ikke har hatt undervisning i emnet. Hvilke misoppfatninger som reduseres, består eller øker, diskuteres før sammendrag, konklusjon og implikasjon.

Erik Holst
Bergen 20. september 2016

Innhold

1 Innledning

Hva skulle oppgaven handle om? Tidlig i prosessen var jeg innom mange ulike emner og det tok rett og slett litt tid å velge et emne til masteroppgaven. Jeg endte opp med misoppfatningen innen betinget sannsynlighet etter grundige overveielser og overtalelse fra min overbevisende veileder. Hva er en misoppfatning? Hvorfor synes elevene at sannsynlighetsregning er så vanskelig? Spørsmålene dukket opp – og kanskje kan jeg finne svar? Hvert enkelt kapittel vil vise avsnittene på første nivå, slik som her:

1.1 Valg av emne og oppgaven

Opp igjennom årene som lærer i den videregående skole har jeg samlet erfaringer innen sannsynlighetsregning og statistikk, først gjennom utdanning og senere som lærer i ulike undervisningssituasjoner. I skolen møter vi utsagn fra elever som (egen samling sitater):

> *«Jeg fikser det meste innen matematikk, men ikke sannsynlighet og statistikk.»*
> *«Jeg skjønner det meste, bortsett fra sannsynlighetsregning. Der skjønner jeg ingenting. Men det burde jo være lett. Tallene er ikke vanskelige.»*
> *«Hvorfor er det så vanskelig med sannsynlighet?»*

Sannsynlighetsregning er karakterisert av situasjoner der resultatet er uforutsigbart og tilfeldig. I motsetning til andre områder i matematikk, er det ofte mye vanskeligere å få bekreftet sine antagelser i situasjoner som involverer sannsynlighet. Det er for eksempel lett å vise at arealet til en sirkel ikke dobles når vi dobler radius, men er det like lett å vise at sjansen for å få en sekser ikke dobles dersom vi dobler antall kast med en terning (side 123 i Dooren, 2014)?

De fleste i oppgavene i sannsynlighet på vg1 er av en slik art at man kan tegne eller telle for å ha kontroll. Sannsynlighetsregning på dette nivået burde kanskje ikke virket avskrekkende, men det gjør det. Det er mange oppgaver som elever helt uten erfaring fra undervisning klarer å løse ved hjelp av egen erfaring, men når tvilen dukker opp kan de lene seg mot egne erfaringer og regler som de tror virker. Sannsynlighet har den egenerfarte virkeligheten både som en støttespiller og som en motspiller. Det er motspilleren vi skal diskutere i denne oppgaven. Vi kan trekke feilaktige slutninger ut fra egen erfaring, uten å se at denne erfaringen strider mot det andre har erfart motsatt. Man er altså seg selv nærmest, også i spørsmål om sannsynlighet.

Oppgaven går ut på å avsløre misoppfatninger som elever i den videregående skole har innen emnet betinget sannsynlighet. Misoppfatninger og prestasjon, eller riktige svar, henger sammen. Derfor er prestasjonen, eller graden av riktige svar, også vesentlig i denne masteroppgaven. Jeg har brukt ordene feilslutning og misoppfatning om hverandre, men ordet misoppfatning forekommer oftest. Begge ordene er definert senere. Temaet for oppgaven er:

Misoppfatninger innen betinget sannsynlighet for elever i den videregående skole.

Temaet omformuleres til problemstilling med tilhørende forskningsspørsmål:

Hvilken innvirkning har undervisning i skolen på misoppfatninger innen betinget sannsynlighet?
Vil undervisning redusere antall misoppfatninger?
Hvilke misoppfatninger består, øker eller minker etter undervisning?

Det kan fort dukke opp andre spørsmål i forbindelse med oppgaven. Det blir i så fall drøftet i kapittel 5. Jeg vil allerede nå presisere at denne masteroppgaven begrenser seg til elever i den videregående skole på studiespesialiserende studieprogram. Elevene som er med i mine undersøkelser er relativt flinke elever på sentrumsskoler i Bergen og har fagene matematikk 1T på vg1 eller S1 på vg2.

Vi vet at det er gjennomført mange undersøkelser om misoppfatninger innen sannsynlighet på ulike trinn i mange forskjellige land. Samtidig er det mindre forskning av eldre dato av sannsynlighet i skolen på grunn av manglende utbredelse av faget. Undersøkelsene viser et sprikende resultat relatert til mitt forskningsspørsmål. I noen emner innen sannsynlighet viser forskning at undervisning reduserer antall misoppfatninger og at prestasjonen forbedres. I andre emner består misoppfatningen, selv etter at undervisning er gitt, og i enkelte emner viser forskning sågar at graden av misoppfatninger øker. Til emnet betinget sannsynlighet har jeg ikke funnet andre undersøkelser som tar for seg prestasjoner til elever før og etter undervisning på vg1 og vg2 nivå, eller tilsvarende.

Til slutt i oppgaven diskuterer jeg implikasjoner svarene kan gi. Kan antallet misoppfatninger reduseres? Kan elever i større grad oppleve mestring innen sannsynlighetsregning dersom spesifikke endringer og tiltak gjøres med læreplanen, utdannelsen og undervisningen til lærerne? Det er viktig å få svar, fordi misoppfatninger innen sannsynlighet er utbredt blant elever på alle nivå.

1.2 Sannsynlighet

At sannsynligheter kan volde problemer for elever og andre er mange forskere enige om:

> «Sannsynlighet kan være et sleipt konsept. Det som gjør sannsynlighet så vanskelig er det store området som emnet dekker. Vi prøver å avgrense en usikkerhet mellom de to ekstreme ytterkantene total ignorering og perfekt kunnskap. (side 139 i Konold, 1991. Min oversettelse)»

1.2.1 Et utgangspunkt, to aspekter og tre teorier

Sannsynlighet er utgangspunktet. Sannsynlighet er formet av to aspekter som står i kontrast til hverandre. Sannsynlighet har et *matematisk aspekt* og et *filosofisk aspekt* – en dualitet. Det matematiske aspektet er konsistent og det hersker bred enighet om dette. Kort fortalt er sannsynlighet et mål på muligheten for at en hending skal skje. Denne sannsynligheten er et bestemt tall mellom 0 og 1, slik at 0 er sannsynligheten for hendingen ikke skjer, og 1 er sannsynligheten for at hendingen helt sikkert kommer til å skje. Det filosofiske aspektet er mye mer omdiskutert og med mange divergerende meninger og tanker. Det kan være designet for å danne en noenlunde grad av tro på at noe spesielt skal skje, uten at vi har tall som kan bekrefte denne troen (side 3 i Chernoff og Russell, 2014). Utledet fra disse to aspektene er det i dag tre regjerende tolkninger av, eller retninger om, sannsynlighet (side 7 i Borovcnik og Kapadia, 2014). Dette støttes også av andre forskere (side 142 Konold, 1991).

- **Klassisk a priori teori** (APT - A Priori Theory, matematisk aspekt):
  a priori, symmetrisk, klassisk tolkning eller teoretisk
- **Frekvensteori** (FQT - Frequentist Theory, matematisk aspekt):
  a posteriori, eksperimentell, frekvens tolkning, empirisk eller objektiv
- **Subjektiv teori** (SJT – Subjective Theory, filosofisk aspekt):
  Bayesisk, intuitiv, personlig, individuell, tro-basert, epistemologisk

I den videregående skole i Norge er det de to første retningene, altså det matematiske aspektet, som er grunnlaget for læreplanene. Dette gjelder også internasjonalt (side 101 i Pfannkuch og Ziedins, 2014). Men dette bildet er på vei til å endre seg noe. Subjektiv sannsynlighet har i større grad fått innpass i læreplaner i mange land.

Klassisk a priori teori: sannsynlighet til en hending er basert på analyse og konstruksjon av utfallsrommet, og bruker i stor grad symmetri, tall, eller enkle geometriske mål, til å spesifisere sannsynligheten for hendingen. En definisjon av sannsynlighet gitt av Abraham de Moivre, ble senere modifisert. Vi kjenner denne definisjonen i dag som Laplace-regelen; gunstige utfall dividert med mulige utfall. Definisjonen er ikke uten problemer. Å finne forholdet mellom gunstige og mulige kombinasjoner er kun riktig dersom vi har en uniform sannsynlighetsmodell. Dette skjer veldig sjelden, bortsett fra i spill med kort og terninger (side 112 i Batanero og Diaz, 2007). Sannsynligheten er gitt «a priori», før og uavhengig av kommende observasjoner og stammer fra en generell lov, for eksempel den fysiske utformingen av en terning.

Frekvensteori: sannsynligheten til en hending baserer seg på opptelling via eksperimentering og simulering, til relativ frekvens i det lange løp. Bernoulli foreslo å finne sannsynligheten til hendinger i hverdagen, ulik hendinger i spill og sjanser, gjennom en tilnærming via relativ frekvens og de store talls lov. Bernoulli beviste at den relative frekvensen i det lange løp nærmet seg den klassiske sannsynligheten for samme hending, og tolket dette som en bekreftelse på at sannsynlighet var et objektivt mål på tilfeldige hendinger (side 112 i Batanero og Diaz, 2007). Sannsynligheten er gitt «a posteriori», etter og avhengig av observasjoner, og er i prinsippet uavhengig av terningens utforming. Hvem har sagt at en terning alltid er rettferdig?

Subjektiv teori: Bayes formel gir mulighet til å finne sannsynligheten for ulike hendinger (årsak) ved observasjon av konsekvensen (effekt) til hendingene. Dersom effekten revideres må sannsynligheten for årsaken endres. Sannsynligheten mister sin objektivitet, som er postulert i de to ovennevnte teorier. Sannsynlighet ble i den sammenheng beskrevet som grad av tro basert på personlige vurderinger og informasjon om eksperimenter relatert til et bestemt utfall. Sannsynligheten til en hending er alltid relatert til et bestemt system av kunnskap og er nødvendigvis ikke den samme for alle. Dette er også vanskeligheten med subjektiv tilnærming, at det er vanskelig å utlede et matematisk uttrykk for sannsynlighet fra en personlig overbevisning. Troen eller tiltroen til en hending er et måleinstrument for sannsynligheten (side 115 i Batanero og Diaz, 2007).

1.2.2 Tilbake- og fremoverblikk

I forhold til andre emner innen matematikk, må vi kunne si at sannsynlighetsregningen har en kort historie. Beskrivelse av sjanse er blitt funnet i gamle kulturer hos indere, babylonere og egyptere, der selv et enkelt terningkast for flere tusen år siden kunne avgjøre mellom liv og død. Dette var en guds vilje, og i tiden for kristendommen var ikke dette et emne klart til å forske på. Et annet moment i denne sammenheng er at sjanse var sjanse – det var ingen vits i å tenke videre på noe som man overhodet ikke kunne forutsi. Den systematiske tilnærmingen begynte ikke før noen tusen år etter de tidligste funn (side 8 i Borovcnik og Kapadia, 2014).

Vår nåværende forståelse av sannsynlighet tok form rundt 1660, En av de mest kjente historiene innen sannsynlighetsregning er den om de to franske herrene Pierre de Fermat (1601 -

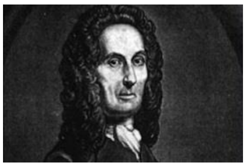
1665) og Blaise Pascal (1623 - 1662) og deres diskusjoner om vinnersjanser i spill. Fermat og Pascal diskuterte og løste to spesifikke problemer; «De Meres Problem» og «Division of Stakes». Det siste problemet beskriver løsninger til fordeling av innsatsen i spill som ikke ble avsluttet (Devlin, 2008). Nederlenderen Christian Huygens (1629 - 1695) ga i 1657 en innføring i sannsynlighetsregning og innarbeidet løsningen til «Division of Stakes» i sin bok «De ratiociniis in ludo aleae». Dette regnes som den første vitenskapelige tilnærmingen til emnet. Senere fulgte bøker fra Jakob Bernoulli (1654/55 - 1705) og Abraham de Moivre (1667 - 1754).


Pierre de Fermat


Blaise Pascal


Christian Huygens


Jakob Bernoulli


Abraham de Moivre

I dag er sannsynlighetsregning og statistikk viktige fag innen forskning, finans og politikk. Mange beslutninger i de ulike bransjene er basert på risikoanalyse, og dermed sannsynlighetsregning og statistikk. I helsevesenet har en økende innsikt i biologiske mekanismer nødvendigvis ikke gjort det enklere å forstå sykdom, identifisere risikofaktorer eller utvikle nye behandlingsstrategier. Økende innsikt gir større kompleksitet, og når kompleksiteten øker blir statistikk et uunnværlig verktøy i medisinsk forskning. Sannsynlighetsregning og statistikk er en del av den kliniske hverdagen i helsevesenet (1424-Skovlund, 2015).

Viktigheten til faget har gjort at det stadig har fått større plass og oppmerksomhet i skolen og høyere utdanningsinstitusjoner, nasjonalt og internasjonalt. Elever møter sannsynlighet i mange situasjoner utenom skolen i sosiale aktiviteter som spill og sport. Som en konsekvens av denne stadige påvirkningen av sjanser og tilfeldigheter i hverdagen, har sannsynlighet blitt en del av skolehverdagen i mange land, også blant unge barn.

Eksempel

Bilde. Lottomillionær fra Verdal.

Statistikk og sannsynlighet har vært en offisiell del av læreplaner i matematikk i engelskspråklige land og Norge siden om lag 1990. Viktigheten av evnen til å forstå og kritisk evaluere statistiske resultat som innvirker på vårt daglige liv, i sosiale og private sammenhenger, har vært påpekt lenge (Wallman, 1993, side 1 i Watson, 2006). Voksne trenger kunnskap for å være i stand til å:

1) tolke og kritisk vurdere statistisk informasjon, data-relaterte argumenter og dokumenter eller stokastiske fenomener som eksisterer i ulike relevante situasjoner.
2) diskutere og kommunisere reaksjoner til statistisk informasjon, som forståelse og mening av denne informasjonen, deres meninger omkring påvirkning informasjonen kan ha, og bekymringer om konklusjoner som informasjonen kan lede til.

Jeg er av den oppfatning at man sannelig trenger å være voksen for å forstå teksten i disse to punktene til Wallman også. Uansett – dette ble også fulgt opp av Watson (1997) som foreslo en tretrinns kunnskapsrakett: Å forstå en tekst – lage en mening av denne – bruke meningen i en sosial kontekst (side 2 i Watson, 2006).

Flere taler for en større utbredelse og sterkere begrepsforståelse av sannsynlighet i skolen. Inntil starten på 90-tallet var det liten eller ingen mulighet for elever å få undervisning og tilegne seg kunnskap rundt begreper som betinget sannsynlighet og uavhengighet (Shaughnessy (1992) side 217 i Tarr og Lannin, 2005). Dette har forbedret seg med nye læreplaner i mange land.

1.2.3 Definisjoner og begreper – og litt engelsk

Begrepsapparatet i sannsynlighetsregning vil gi utfordringer for de fleste elever på et eller annet nivå. Det kan være en utfordring å holde oversikten over de norske og de engelske begrepene. Jeg vil kort ta for meg noen av de vanligste begrepene som elevene møter i læreplanen for ungdomsskolen og første trinn på vg1 - 1T (Emnekode MAT1-04), og i tillegg se på noen engelske varianter av disse begrepene.

1.2.3.1 Læreplaner

Emner som er relatert til sannsynlighet og statistikk og anvendelsen av disse, har fått større plass i læreplanene de senere år. Elevene starter med sannsynlighet og statistikk allerede på barneskolen.

Statistikk, sannsyn og kombinatorikk i læreplanen for ungdomsskolen og studiespesialiserende vg1:

Statistikk omfattar å planleggje, samle inn, organisere, analysere og presentere data. I analysen av data høyrer det med å beskrive generelle trekk ved datamaterialet. Å vurdere og sjå kritisk på konklusjonar og framstilling av data er ein sentral del av denne prosessen. I sannsynsrekning talfester ein kor stor sjanse det er for at ei hending skal skje. I kombinatorikk arbeider ein med systematiske måtar for å telje opp moglege utfall for å kunne berekne sannsyn.

Mål for opplæringa for 8., 9. og 10. trinn

- gjennomføre undersøkingar og bruke databasar til å søkje etter og analysere statistiske data og vise kjeldekritikk
- ordne og gruppere data, finne og drøfte median, typetal, gjennomsnitt og variasjonsbreidd, presentere data, med og utan digitale verktøy, og drøfte ulike dataframstillingar og kva inntrykk dei kan gje
- finne og diskutere sannsyn gjennom eksperimentering, simulering og berekning i daglegdagse samanhengar og spel
- beskrive utfallsrom og uttrykkje sannsyn som brøk, prosent og desimaltal
- drøfte og løyse enkle kombinatoriske problem

Mål for opplæringa for vg1 – 1T

- formulere, eksperimentere med og drøfte uniforme og ikkje-uniforme sannsynsmodellar
- berekne sannsyn ved å telje opp gunstige og moglege utfall, systematisere oppteljingar ved hjelp av krysstabellar, venndiagram og val-tre og bruke addisjonssetninga og produktsetninga

Kompetansemålene for vg1 – 1P varierer noe fra 1T, men har stort sett samme fokus. Læreplanene for S1 og R1 er i appendiks.

1.2.3.2 Noen av begrepene i læreplanen

Vi skal se på noen begreper i læreplanen for ungdomsskolen og for den videregående skole. Allerede på ungdomsskolen støter elevene på ordet kombinatorikk. Å kunne drøfte og løse enkle kombinatoriske problemer vil si å kunne bruke multiplikasjonsprinsippet og kunne utarbeide enkle, men fullstendige utfallsrom og å finne sannsynligheten til disse ved bruk av addisjonssetningen og produktsetningen med valgtre. Å kunne utarbeide fullstendige utfallsrom er en viktig forutsetning for å kunne svare riktig på kombinatoriske spørsmål.

En begivenhet, hendelse eller mitt foretrukne begrep; hending, er omtalt i hovedmålet i læreplanen i forbindelse med at sannsynlighet er en tallfesting av sjansen for at en hending skal skje, noe som harmonerer med det matematiske aspektet (kapittel 1.3.1). Begrepet hending er ellers lite brukt i læreplanene, i motsetning til engelskspråklige læreplaner og bøker som gir ordet hending (event) en mye større og sentral plass. Engelskspråklige bøker og artikler bruker begreper som «Simple Events», «Compound Events» og «Combined Events». En enkel hending (Simple Event) er knyttet til et enkelt forsøk. Det er vanlig i norske bøker å si at en hending består av ingen, ett eller flere utfall (side 109 i Ekern med flere, 2004b). Et terningkast med hendingen: «A: Antall øyne er partall» er derfor et eksempel på en enkel hending (Simple Event). Hendingene $A \cup B$ og $A \cap B$ er eksempler på sammensatte hendinger, på engelsk «Compound Events» (side 191 i Polaki, 2005). To andre engelske ord som bedre beskriver snittet mellom to hendinger er «intersection», og «conjunction». Ordet «conjunction» tilsvarer i sannsynlighetsregning det norske ordet «og», som igjen tilhører ordklassen konjunksjoner. To hendinger A og B, og deres snitt, blir på engelsk: «Two events and their conjunction». Jeg vektlegger ordet «conjunction» fordi en senere omtalt misoppfatning knyttes til ordet.

Under det siste punktet i læreplanen for vg1T er begrepene *krysstabell*, *venndiagram* og *valgtre* nevnt, samt at elevene skal bruke *addisjonssetningen* og *produktsetningen*. Krysstabell, venndiagram og valgtre er hjelpemidler ved bruk av reglene addisjonssetningen og produktsetningen. I den sammensatte hendingen $A \cup B$ får vi ofte bruk for addisjonssetningen, eventuelt sammen med et venndiagram eller en krysstabell. Den sammensatte hendingen $A \cap B$ kan være et to-dimensjonalt forsøk (synkrone hendinger og krysstabell) eller to suksessive hendinger A og B (asynkrone hendinger og valgtre). To synkrone hendinger skjer samtidig, i motsetning til asynkrone hendinger. Dette beskrives grundigere i kapittel 2.4.

Å forstå sannsynligheter som involverer sammensatte hendinger vil si å (side 192 i Polaki, 2005):

a) Kunne lage et komplett utfallsrom for hvert eksperiment.
b) Kunne bruke symmetri og eksperimenter til å lage en sannsynlighetsfordeling.
c) Kunne bruke a og b i en kompleks sammensetning med kontekst.

Dette er vesentlig, men er elevene i stand til dette – og eventuelt – hvorfor ikke?

1.3 Forskning

Forskning innen tenkning i sannsynlighetsregning er av relativ ny dato, men har økt stort i omfang de siste 60 år. Det har resultert i viktige bidrag fra blant annet Piaget og Inhelder (1951/1975), Fishbein (1975), Shulte og Smart (1981), Hawkins og Kapadia (1984), Garfield og Ahlgren (1988), Kapadia og Borovcnik (1991/2009), Shaughnessy (1992/2003), Borovcnik og Pears (1996), Jones (2005), Burrill og Elliot (2006), Jones, Langrall og Mooney (2007) og Biehler og Pratt (2012). Forskningen fra disse personene er delt inn i tre perioder av Jones og Thornton (2005):

- Piaget-perioden; definert av Piaget og Inhelder (50- og 60-tallet).
- Etter Piaget-perioden; definert av Fishbein, Tversky og Kahneman (70- og 80-tallet).
- Samtidsforskningsperiode; se nedenfor (90- og 2000-tallet).

Den siste perioden kjennetegnes ved forskning på undervisning og læring i klasserommet, som en konsekvens av at sannsynlighetsregning har fått større utbredelse i læreplaner i skoler verden over (Chernoff og Sriraman, 2014).

Den senere tids økning i arbeidsgrupper og diskusjonsgrupper på større utdanningsmesser verden over, spesielle tidsskrifter og dedikerte kapitler i håndbøker er indikatorer på at feltet sannsynlighet er på vei fremover. Chernoff og Maririnam foreslår å kalle den neste perioden i forskningen innen sannsynlighetsregning og statistikk for «Assimilation Period». Da tenkning innen sannsynlighet får større og større plass i matematikkundervisningen, bør den nye perioden tilpasses dette.

Ved nøye studier og sondering av studenters ideer om betinget sannsynlighet dukker det opp noen misoppfatninger og feilslutninger. Disse misoppfatningene eller feilslutningene krever å bli studert nøye, for å få en nødvendig oppklaring (side 292 i Falk, 1986).

Man kan spørre seg om mennesker har evnen til å være intuitiv statistiker, se rent på sannsynligheter, eller om vi er forutinntatt og tilbøyelig til å tenke feil, når vi resonnerer rundt sannsynlighet. Forskning på feltet i 50- og 60- årene viste en majoritet mot det første – intuitive statistikere, mens forskning på 70- og 80-årene i all hovedsak konkluderte med det andre – forutinntatt og tilbøyelig til å tenke feil. For å studere kløften mellom disse to periodene nærmere, har psykologer startet arbeidet med å identifisere og karakterisere under hvilke omstendigheter folk, både barn og voksne, er i stand til å resonnere fornuftig og riktig. Blant annet har Tversky og Kahneman påstått at menneskets hjerne ikke er laget til å tenke på sannsynlighet (side 128 i Meder og Gigerenzer, 2014). Det må med andre ord læres.

Forskningen har i mindre grad sett på hvordan undervisning påvirker misoppfatningene i sannsynlighet. De første som spesifikt rettet forskning mot påvirkningen undervisning kan ha på misoppfatninger innen betinget sannsynlighet var Fishbein og Gazit (1984). Dette ble gjort på grunn av dårlige prestasjoner blant elever som hadde fulgt et instruksjonsprogram. Prestasjoner før og etter undervisning er et vesentlig moment i denne masteroppgaven.

1.4 Betinget sannsynlighet

Betinget sannsynlighet spiller en sentral rolle i prosessen med å kunne si noe relativt sikkert om en usikker verden (side 292 i Falk, 1986). Betinget sannsynlighet $P(B|A)$ er også et eksempel på en sammensatt hending, men bruker som oftest sin egen term, altså «betinget sannsynlighet». Dette vil vi se nærmere på i dette avsnittet. Vi kan ikke snakke om *betinget sannsynlighet* uten å kjenne til *avhengige hendinger*. Det er verdt å merke seg at ingen av disse begrepene er nevnt eksplisitt i læreplanene, selv om begrepene absolutt ligger implisitt i begrepet *produktsetningen*. Vi kan si at

uavhengighet er et spesialtilfelle av avhengighet og dermed et spesialtilfelle av betinget sannsynlighet. I norske lærebøker er dette en vanlig definisjon av betinget sannsynlighet:
I et forsøk har vi to hendinger A og B. Definisjonen av betinget sannsynlighet for hendingen B|A er da (side 81 i Ekern med flere, 2004a):

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

I noen eksperimenter med tilfeldig utfall er det kun de utfallene som er elementer i delmengden A i utfallsrommet U som er interessante. Under disse omstendigheter er den betingede sannsynligheten for en hending B, gitt at A har hendt, kun de elementer av B som også er i A. Det vil si at sannsynligheten for B er beregnet under nye forutsetninger, eller et nytt utfallsrom gitt av betingelsen A (side 215 i Tarr og Lannin, 2005). Med noen enkle algebraiske operasjoner utført på definisjonen ovenfor, får vi produktsetningen for to avhengige hendinger A og B:

$$P(A\cap B)=P(A)\cdot P(B|A)$$

I eksperimenter med en serie med forsøk får vi bruk for produktsetningen. Definisjonen av betinget sannsynlighet og produktsetningen er vesentlig i denne masteroppgaven.

1.5 Misoppfatninger

I dette avsnittet ser jeg nærmere på misoppfatninger generelt og begreper og definisjoner knyttet til dette. På slutten av kapittelet vil jeg se spesielt på en misoppfatning som ikke er direkte knyttet til sannsynlighet, men som likevel spiller en viktig rolle – også innen sannsynlighet. Misoppfatninger innen sannsynlighet er plassert i kapittel 2.

1.5.1 Definisjoner og begreper – og litt engelsk igjen

Ufullstendige tanker knyttet til et begrep kaller vi for misoppfatninger (side 14 i Brekke, 2002). Misoppfatning skiller seg fra feil, ved at en feil kan forekomme tilfeldig. En misoppfatning er altså ikke tilfeldig. Bak en misoppfatning ligger det en bestemt tanke eller ide, som det ofte kan være vanskelig å rokke ved. En misoppfatning er ofte et resultat av overgeneralisering av tidligere kunnskap til nye områder, der disse kunnskapene ikke er fullstendig dekkende. Eksemplet nedenfor viser en overgeneralisering fra hele tall til desimaltall.

Eksempel
Dersom noen påstår at 7,8 er større enn 7,9, er det en feil og ikke en misoppfatning. Det er en feil fordi de like gjerne kunne påstått at 7,4 var større enn 7,5.
Dersom noen påstår at 7,10 er større enn 7,9, er det en overgeneralisering fra hele tall om at 10 er større enn 9 – altså en misoppfatning. Muntlig formulering av eksemplet forsterker misoppfatningen, fordi vi mister den visuelle opplevelsen av vårt tallsystem som et posisjonssystem.

Vi kan se på overgeneralisering som et forsøk på å skape mening i en ny sammenheng med gammel kunnskap. Et annet ord som brukes i forbindelse med misoppfatning er forestilling. En forestilling er en forklarende modell for å forstå verden rundt oss. Siden en forestilling kan være både riktig og gal, kan vi ikke kalle en forestilling for en misoppfatning (side 285 i Savard, 2014). Ordet forestilling brukes lite i denne masteroppgaven, men andre begreper som dukker opp er heuristikk (heuristisk skjevhet) og feilslutning. Litt norsk - engelsk:

- *Misoppfatning*; misconception, misperception, misapprehension
- *Feil;* error, fault, inaccuracy, miscalculation, misunderstanding, misinterpretation
- *Feilslutning*; fallacy
- *Heuristikk;* heuristic (psychological bias)

I mangel av en bedre egen beskrivelse av heuristikk, annet at det litt enkelt sagt er synonymt med tommelfingerregel, velger jeg å ty til Store Norske Leksikon (https://snl.no/heuristikk%2Fpsykologi):

Heuristikk, enkel fremgangsmåte eller strategi som en problemløser kan ta i bruk for å øke sjansen til å løse en oppgave. Eksempler er å se etter analogier, arbeide baklengs, eller tenke i middel-mål-relasjoner. Beslutninger og bedømmelser i dagliglivet baserer seg for en stor grad på heuristikker, som f.eks. det å velge det kjente fremfor det ukjente, eller å foreta sammenligninger med en «forankrings-verdi». Slike bedømmelser er effektive, men garanterer ikke riktig resultat og kan dermed føre til utfall som har systematiske skjevheter, bias.

Vi kan diskutere om mennesker overhodet kan forholde seg nøytral til sannsynlighetsregning. Jeg vil si at vi alle er sannsynlighetstenkere i mer eller mindre grad i nesten hver eneste avgjørelse vi må ta. Med dette utgangspunktet er det soleklart at det finnes systematiske skjevheter i våre resonnementer.

1.5.2 Elevvansker

Veldig ofte vil ikke elever ha et nøytralt forhold til problemer innen sannsynlighetsregning. En god del forskning viser at personer tar i bruk et mindre sett av heuristiske forestillinger i prosesser eller beregninger om sannsynlighet. Slike heuristikker resulterer ofte i rask og generell tenkning rundt sannsynlighet og kan lede til konklusjoner som strider mot teorien. (side 393 i Konold et al, 1993). De mulige utfallene kan ha en personlig forbindelse til eleven, emosjonelt og materielt. Hans Rosling viste i et enkelt eksperiment at et nøytralt forhold til et problem kan gi et bedre resultat enn om man har et personlig forhold til problemet. Spørsmålet er hvor mange barn kvinner i Bangladesh føder i gjennomsnitt. Bildet viser at 12 prosent av de spurte svarte riktig. Med fire mulige utfall påstår Rosling at en uvitende sjimpanse ville svart riktig i 25 prosent av tilfellene.

Bilde. Hans Rosling: Dokumentar om jordens befolkning. Kilde: NRK.

Lenge før elever har undervisning i sannsynlighet har de vært i et uendelig antall situasjoner som involverer usikkerhet. De har en forståelse av når de ulike ordene skal brukes i daglige situasjoner, men dette lar seg ofte ikke overføre inn i en formell kontekst med teori om sannsynlighet. Dette leder til misforståelser, forvirring og kollaps i kommunikasjonen med seg selv. Dersom læreren klarer å se studentenes ståsted og tolkning av sannsynlighet, kan læreren bli bedre i stand til å lede eleven i riktig retning (side 144 i Konold, 1991). Fra et konstruktivistisk ståsted kan vi si at lærerens rolle er å hjelpe elevene til å konstruere sin egen kunnskap gjennom bekjemping av heuristiske skjevheter. Men hvordan gjør vi det?

1.5.3 Diagnostiske oppgaver

Vi kan bruke ulike verktøy for å avgjøre hvilke misoppfatninger en elev har. En metode er intervju, som er nærmere beskrevet i kapittel 3. En annen metode er å bruke diagnostiske oppgaver, gjerne sammen med intervju. I denne masteroppgaven er diagnostiske oppgaver sentrale og jeg vil derfor først si noe generelt om diagnostiske oppgaver og hva som kjennetegner disse.

1.5.3.1 Hva er og når brukes diagnostiske oppgaver?

Diagnostiske oppgaver er ikke oppgaver som forekommer på en prøve, selv om en diagnostisk oppgave kan være identisk til en oppgave på en prøve. Diagnostiske oppgaver kommer, som vi også skal se i denne masteroppgaven, ofte før en undervisningssekvens i et bestemt emne, eller etter en undervisningssekvens. Diagnostiske oppgaver blir brukt til (side 15 i Brekke, 2002):

- å identifisere og framheve misoppfatninger som elevene har utviklet
- å gi læreren informasjon om løsningsstrategier elevene bruker for ulike typer av oppgaver
- å rette undervisningen mot å framheve misoppfatningene
- å utvikle elevenes eksisterende løsningsstrategier
- å måle hvordan undervisningen har hjulpet elevene til å overvinne misoppfatningene

Tillegget til det siste kulepunktet sier at dette måles ved å bruke de samme oppgavene både før og etter undervisningssekvensen. Alle kulepunktene ovenfor griper inn i forskningsspørsmålet til denne masteroppgaven. Diagnostiske prøver eller tester kan gjerne inneholde noen typer oppgaver som elevene ikke tidligere har arbeidet mye med. Elevene vil likevel oftest ha ideer om hvordan de skal angripe oppgavene for å finne et svar. Testen i denne masteroppgaven tilfredsstiller i stor grad disse kravene.

1.5.3.2 Konstruksjon av diagnostisk oppgave

Enkelte lærere kan føle at det er feil eller «urettferdig» å prøve elevene i lærestoff som ikke er gjennomgått. Det bør gjøres klart for elevene at oppgavene blir brukt på en annen måte enn det elevene er vant med. Hovedmålet med en diagnostisk test er (side 16 i Brekke, 2002):

- å oppdage hvilke tanker de har om ulike begreper,
- å bli kjent med de vanskene som er knyttet til disse begrepene,
- å hjelpe læreren med å planlegge undervisningen.

Oppgavene har ikke til hensikt å vurdere eleven med tanke på rangering. I en diagnostisk oppgave prøver vi å unngå å stille spørsmål der elevene kan få riktige svar selv om de har feilaktige ideer knyttet til begrepet. Det er altså mulig å skille mellom oppgaver ut fra den diagnostiske informasjonen de gir om hvilke tanker elevene gjør seg om et bestemt begrep. Når man står overfor diagnostiske oppgaver, har elever liten eller ingen forståelse hvordan slike oppgaver skal beregnes, og noen tror at problemet i seg selv ikke er løsbart fordi det mangler rot i virkeligheten (Falk 1994, side 6 i Tomlinson og Quinn, 1997). I kapittel 4 møter vi ulike diagnostiske oppgaver.

1.5.4 Misoppfatning eller bare feil?

Vi møter nå igjen ordet «feil» i forbindelse med misoppfatninger. I innledningen skrev jeg at en misoppfatning skiller seg fra feil, ved at en feil kan forekomme tilfeldig. Å si at en feil er tilfeldig er kanskje å se for enkelt på saken. Dersom et svar på en oppgave er feil, har da eleven misforstått oppgaven? Svaret på dette er ikke entydig ja. Eleven kan ha gått lei og skriver omtrent det første og beste som faller ham inn. Når dette skjer, kan vi ha en såkalt «system 1»-feil (Kahneman, side 6 i Shleifer, 2012). Våre resonneringssystemer kan deles inn i to:

- System 1: den automatiske og ubevisste prosessen som vi bruker hele tiden.
- System 2: den bevisste, analyserende og resonnerende prosessen som vi bruker sjeldnere.

Selv om vi ønsker at system 2 skal være det herskende resonneringssystem, at vi er på hele tiden, må vi ikke se på system 1 som underlegent system 2. System 1 er effektivt og tidsbesparende, men også tilbøyelig for feilslutninger (1469-Brakedal, 2015). Denne dual-prosessen, vekslingen mellom system 1 og system 2 som resonneringssystem, er viktig. Det er viktig å kunne veksle mellom det automatiske og det kontrollerte. Delingen mellom system 1 og system 2 minimaliserer innsatsen og optimaliserer utbyttet (Kahneman, side 7 i Shleifer, 2012).

Eksempel

Vi kan tenke oss to følgende situasjoner; en bilist og en fotballspiller:
Begge må ha mange oppgaver som er på autopilot - system 1, for å kunne konsentrere seg om de viktige oppgavene, og dermed koble inn system 2, i vanskelige situasjoner.
En dyktig bilist reagerer raskt og fornuftig i utfordrende situasjoner. En bilist som må koble inn system 2 hver gang blinklyset skal settes på, blir neppe dyktig.
En fotballspiller som må tenke seg om både en og to omganger om den aktuelle pasningen skal slås med innsiden eller utsiden av foten, vil neppe fylle tribunene.

System 2 slår inn når situasjonen krever det, men ikke alltid. Viktige valg som er foretatt på system 1 nivå, har mulighet for dramatiske avvik fra det normale, noe som kan lede til brilliante løsninger, men også mange systematiske feil (Kahneman, side 8 i Shleifer, 2012). Kahneman sier at områdene for system 1 og system 2, og dermed grensene mellom de to systemene, varierer fra person til person. Etter hvert som kunnskap tilegnes, vil grenselinjen mellom de to systemene forskyves i favør system 1. Eksemplet nedenfor viser det brede spekteret av system 1-feil.

Eksempel

En tennisracket og en tennisball koster til sammen 110 kr. Racketen koster 100 kr mer enn ballen. Hvor mye koster ballen?

Dersom svaret må avgis temmelig raskt, er det en stor tendens til å svare at ballen koster 10 kroner, også blant studenter på høyere nivå. Ved å bli gitt anledning til å koble inn system 2 vil sannsynligvis flere klare å finne det riktige svaret, 5 kroner.

Eksempelet på neste side viser at overgangen fra system 1 til system 2 også er viktig i helsevesenet.

Eksempel

La oss anta at sannsynligheten for at en pasient har en sykdom er 25 % og at det i tillegg foreligger en test på sykdommen med negativt svar (det vil si at testen viser at pasienten er frisk). Testen viser et negativt resultat i 18 % av tilfellene, selv om personen har sykdommen. Skal legen friskmelde personen på bakgrunn av opplysningene ovenfor?

En lege bør i dette tilfellet ikke gå på autopilot, kun bruke system 1. Da er det stor sjanse for feilslutning. En lege som kobler inn system 2, vil med god anamnese (relevante diagnostiske oppfølgingsspørsmål) redusere sannsynligheten for feilslutninger betraktelig.

Det er viktig å presisere at system 1-feil, ikke kan relateres til noen spesifikk misoppfatning. En heuristikk forbundet til system 1 kan beskrives som den raske prosessen fra informasjon til vurdering. Heuristikkene reduserer større problemer til enklere kognitive oppgaver, og fører til effektiv beslutningstaking. Vurderingene man får ved å bruke heuristikker er ofte gode nok, men kan også føre til systematiske skjevheter. Disse skjevhetene dukker også opp i sannsynlighetsregning.

Ved siden av Kahneman og Tversky er det mange andre forskere (Evans, Levinson, Pollock, Klein, Snyder med flere) som her beskrevet dualiteten i tenkningsprosessen, og brukt ulike begreper for å beskrive system 1 og system 2 tenkningen. For en utvidet oversikt se side 659 i Stanovich og West, 2000.

2 Misoppfatninger innen sannsynlighetsregning

Etter hvert som jeg leste om ulike misoppfatninger i sannsynlighetsregning, økte andelen jeg skrev om dramatisk. Jeg har likevel gjort et forsøk på å begrense antallet misoppfatninger i denne oppgaven. For en mer utfyllende liste over ulike misoppfatninger henviser jeg til kapittel 7.4. Mange misoppfatninger blir listet opp og snakket om, men det som står sentralt i tenkningen rundt betinget sannsynlighet er evnen til å kunne begrense det opprinnelige utfallsrommet etter de gitte betingelsene.

2.1 Misoppfatning 1 Oppgaver med kontekst

I sannsynlighetsregning finner vi mange oppgaver med kontekst. I hvilken grad kontekst i en oppgave er et bidrag eller et hinder for at misoppfatninger oppstår eller fjernes er et viktig spørsmål. Et annet spørsmål er om forvirring er et mer dekkende ord enn misoppfatning, fordi de fleste oppgaver som oftest har en kontekst som det sjelden er interessant å fjerne. Selve konteksten og språket som er brukt kan ha stor innvirkning på prestasjoner til elever, og spesielt i oppgaver med betinget sannsynlighet (side 559 i Lecoutre, 1992). Det er eksempler på at elever forveksler sannsynlighetene $P(A|B)$, $P(\mathrm{B}|\mathrm{A})$ og $P(A\cap B)$ i oppgaver (side 60 i Watson og Moritz, 2002).

Min erfaring er at elever ofte blir stilt ovenfor oppgaver med en uklar tekst som mer eller mindre har til hensikt å lage forvirring. Eksemplet nedenfor viser øverst to tilfeller som elever har problemer med å skille fra hverandre (side 81 i Ekern med flere, 2004a). Nederst er konteksten fjernet. Å fjerne konteksten og definere hendingene A og B som vist, vil gi større grad av riktige svar.

Eksempel **Fjerning av kontekst**

En krysstabell er gitt slik at følgende kan besvares. Finn sannsynligheten for at

a) En tilfeldig valgt sekk er fra parti B og har egenskap A.
b) En tilfeldig valgt sekk fra parti B har egenskap A.

Vi har samme krysstabell, definerer hendingene A og B og spør: Finn:

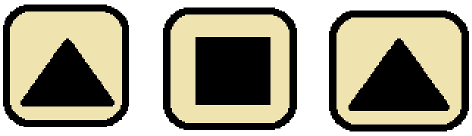
a) $P(B\cap A)$
b) $P(A|B)$

At to forsøk er isomorfe, betyr at ideen bak dem er like, kalkulasjonen og svarene er like, men de har ulike kontekster (side 558 i Lecoutre, 1992). Se eksempelet på neste side (side 562 i Lecoutre, 1992):

Eksempel To ulike kontekster

Forsøk 1 Geometrisk kontekst

Du skal trekke ut to av de tre kortene og lage en sammensatt figur; som enten er et hus eller en rombe (☗ eller ◆).



Spørsmål 1 – Tror du at det er

a) Lik sjanse for hus og rombe?
b) Større sjanse for hus enn rombe?
c) Større sjanse for rombe enn hus?
d) Umulig å svare? Hvorfor?

Forsøk 2 Sjanse kontekst

To drops farget oransje og en var gul.

Du skal trekke to av de tre dropsene

Spørsmål 2 – Tror du at det er

a) Lik sjanse mellom en av hver farge og to med oransje farge?
b) Større sjanse for en oransje og en gul?
c) Større sjanse for to oransje?
d) Umulig å svare? Hvorfor?

Resultat i undersøkelsen med 87 elever i alderen 15 til 17 år:

	a	b	c	d
Spørsmål 1	0,23	0,75	0,02	0,00
Spørsmål 2	0,44	0,45	0,06	0,04

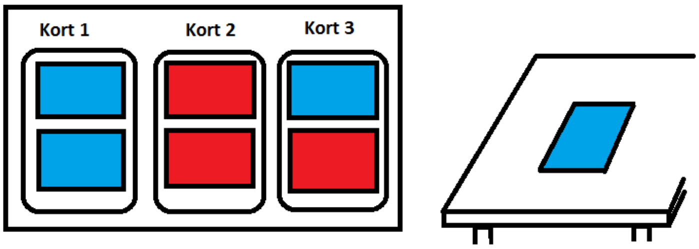
Andelen som anså sjansene for like (alternativ a) var lavere i forsøk 1 enn i forsøk 2, og omvendt for det riktige svaret (alternativ b). Konteksten spiller altså en rolle. Andre funn:

- elevene som svarer at sjansene er like store (alternativ a) tenker at det er en betingelse til stede, at vi må minst ha en trekant eller minst en oransje drops. De trekker denne vekk fra utfallsrommet, slik at det er to elementer igjen, en av hver type.
- elevene som gjør tellefeil, ser at det er flere trekanter og derfor større sjanse for rombe. Det er flere oransje drops og derfor større sjanse for to av dem.

Funnene er konsistente med andre funn i tilsvarende forsøk med isomorfe oppgaver. I noen undersøkelser brukte forskere oppgaver som var spekket med kontekstuelle hindringer (Shaughnessy (1992) side 61 i Watson og Moritz, 2002). Hindringene kan være store fordi den betingede hendingen ligger delvis skjult i teksten. Dette er dokumentert mange ganger. Vi ser på følgende eksempel:

Eksempel

I en boks eller hatt er det tre kort. Ett kort er blått på begge sider, ett er rødt på begge sider og ett kort har en blå og en rød side. Vi trekker ut ett av kortene og legger kortet på bordet. Det er en blå side som vender opp.



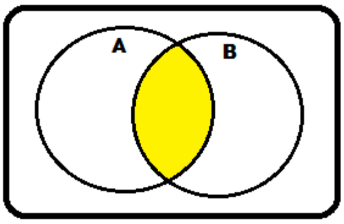
Hvor stor er sannsynligheten for at den siden som vender ned mot bordet også er blå?

Mange vil gjøre feil antagelse om at sjansen er ½ fordi siden som ligger ned også er blå. Hva er betingelsen? Det røde kortet (RR) er ute og altså har vi endret utfallsrommet, og burde spurt på nytt. Sannsynligheten for blått kort ned er 2/3. Vi kan ikke unngå oppgaver med kontekst og det er derfor viktig å være lesekyndig og observant. En annen oppgave som belyser samme sak finner vi i «A tale of two Probabilities» (side 49-52 i Falk og Kendig, 2012), en annen vinkling på betinget sannsynlighet.

Denne masteroppgaven vil ikke svare på forskjellen i prestasjon mellom kontekst eller ikke og heller ikke forskjellen i prestasjon mellom to isomorfe oppgaver, men prøve å gi svar på om undervisning hjelper på prestasjonen, og dermed at forveksling reduseres på de samme oppgavene med kontekst.

2.2 Misoppfatning 2 Snittet mellom to hendinger

La oss se på to ikke disjunkte hendinger A og B. Den sammensatte hendingen $A \cap B$ er det fargelagte området i venndiagrammet nedenfor.



Vi ser ut fra figuren at snittet mellom A og B er mindre enn de to enkelte hendingene A og B. Det tilsvarende vil gjelde for sannsynligheten $P(A \cap B)$ og vi har:

$$P(A \cap B) \leq P(A) \quad \text{og} \quad P(A \cap B) \leq P(B).$$

Mangel på identifikasjon av ulikhetene ovenfor er en feilslutning. Tversky og Kahneman gav denne feilslutningen navnet «conjunction fallacy» (side 294 i Tversky og Kahneman, 1983/side 290 i Savard, 2014).

Feilslutning om snittet mellom to hendinger er neglisjering av regelen

$$P(A \cap B) \leq P(A) \quad \text{eller} \quad P(A \cap B) \leq P(B).$$

En velkjent diagnostisk oppgave om feilslutninger om snitt i sammensatte hendinger er den såkalte «Linda jobber i bank» » (side 59 i Tversky og Kahneman, 1974).

Eksempel **Linda 1**

Dersom Linda var en 31 år gammel singel kvinne som ikke er redd for å ta ordet og er engasjert i samfunnssaker som nedrustning og likeverd, hvilken av påstandene nedenfor er mest sannsynlig?

A: Linda er en bankfunksjonær

B: Linda er en bankfunksjonær og aktiv i feministbevegelsen

Løsning:

Antallet kvinnelige bankfunksjonærer som er aktivister er inkludert A, da er mengden i A nødvendigvis større enn eller lik mengden i B. Vi har altså at $B \subset A$ og dermed at $P(B) \le P(A)$.

Ifølge Kahneman og Tversky svarte mer enn 80 % av de spurte, inkludert flere som var skolert i statistikk og sannsynlighet, alternativ B. (side 2 i Tomlinson og Quinn, 1997). Konteksten i oppgaven er spesiell på mange måter og kjennetegner med dette en diagnostisk oppgave. Den er konstruert for at man skal velge vekk alternativ A. Det rettes kritikk mot de kontekstuelle fellene i Linda-problemet, og andre diagnostiske problemer som forekommer i slike undersøkelser (side 62 i Watson og Moritz, 2002). Denne kritikken blir delvis tilbakevist av andre forskere som arbeidet videre med «Linda-problemet». De omorganiserte oppgaven slik at elevene *ikke* ble ledet til:

- A: å tro at Linda *ikke* var aktiv i feminist-foreningen. Direkte motsetning til B.
- B: å overse det viktige ordet «og». Konjunksjonen skulle være godt synlig.

Denne organiseringen av oppgaven viste seg stort sett å være nytteløst, conjunction fallacy bestod (side 124 i Batanero og Diaz, 2007). Eksemplet er omtalt i mange av artiklene jeg har lest. For eksempel artikkelen av Watson og Moritz (2002).

2.3 Misoppfatning 3 Motsatte betingede sannsynligheter

Vi har to hendinger A og B. Da er $P(B|A)$ og $P(A|B)$ to motsatte betingede sannsynligheter. Det finnes flere ulike misoppfatninger som omhandler to motsatte betingede sannsynligheter.

2.3.1 3A Forveksling og likhet

Det har alltid eksistert en språklig forvirring rundt to omvendte sannsynligheter. I noen tilfeller er $P(B|A) = P(A|B)$, men som oftest ikke. Vi må være i stand til å skille dem fra hverandre. Et kort eksempel nedenfor (side 66 i Watson og Moritz, 2002)

Eksempel

a) Anslå sannsynligheten for at en kvinne er lærer.

b) Anslå sannsynligheten for at en lærer er kvinne.

Løsning:

Det er rimelig å anslå at $P(b) > P(a)$.

I undersøkelsen til Watson og Moritz var det flere som forvekslet de to hendingene. De mente da at det kunne være konteksten som forvirret elevene, men kunne ikke si dette sikkert ovenfor (side 71 i

Watson og Moritz, 2002). Slike forvekslinger forekommer på alle nivå blant elever og lærere (side 295 i Falk, 1986).

Å svare på sannsynligheten $P(A|B)$ i stedet for den omvendte sannsynligheten $P(B|A)$ er en feilslutning om den omvendt betingede sannsynligheten.

Falk kalte også denne misforståelsen «Confusion of the inverse» (side 295 i Falk, 1986) eller «fallacy of the transposed conditional» (Diaz og Fuente, 2007).

Oppgaver som likner eksempelet ovenfor forekommer ofte i en kontekst som omhandler sykdom og testresultater. Falk viser til Eddy (1982) der «sannsynligheten for sykdom (S) gitt positivt utslag (U)» feilaktig blir likestilt med «sannsynligheten for positivt utslag (U) gitt sykdom (S)», altså $P(\mathrm{S}|\mathrm{U})$ blir likestilt med $P(\mathrm{U}|\mathrm{S})$. I et eksempel viser Falk til Pauker og Pauker (1979) hvilke dramatiske forskjeller slike forvekslinger kan ha (side 295 i Falk, 1986).

Eksempel
Kvinner på 30 år har en sannsynlighet 1/885 for å føde et barn med Downs syndrom, altså $P(D)=0,00113$. Dersom den betingede sannsynligheten i en test på Downs syndrom viste riktig i 99,5 % av tilfellene, enten fosteret hadde syndromet eller ikke, ville vi få $P(POS\,|\,D)=0,995$ og $P(\mathrm{NEG}\,|\,\mathrm{ikke}\,D)=0,995$. Den motsatte betingede sannsynligheten til $P(POS\,|\,D)$ er $P(\mathrm{D}\,|\,\mathrm{POS})$. Utregningen for denne sannsynligheten gir:

$$P(\mathrm{D}\,|\,\mathrm{POS})=\frac{P(D\cap POS)}{P(POS)}=\frac{P(D)\cdot P(POS\,|\,D)}{P(POS\,|\,D)+P(POS\,|\,ikke\,D)}$$
$$=\frac{0,00113\cdot 0,995}{0,995\cdot 0,00113+0,005\cdot(1-0,00113)}=0,184$$

Det er altså ingen grunn til å ta lett på dette. Når det gjelder to omvendt betingede sannsynligheter må man stille seg minst tre spørsmål:

- Tror elevene at to omvendt betingede sannsynligheter er to sider av samme sak – og dermed like?
- Tror elevene at to omvendt betingede sannsynligheter er komplementære?
- Oppfatter elevene forskjellen, men forveksler dem likevel?

Selv om spørsmålene er stilt, blir de ikke besvart i samme detaljerte grad. Jeg vil kun avgjøre om elevene forveksler to omvendt betingede sannsynligheter, ikke studere årsaken.

2.3.2 3B Inversjon av tidsaksen

Kausalitet er evnen til å forårsake en effekt, relasjonen mellom en hending A (årsaken) og en annen hending B (effekten), hvor det menes at A er en nødvendig betingelse for B. De fleste elever skjønner og aksepterer det faktum at utfallet til en hending A kan virke inn på utfallet til en senere hending B. De sier samtidig at utfallet til hending B ikke kan påvirke utfallet til en hending A, som da allerede har skjedd. Det er en utbredt feilslutning at betingelser må komme på riktig tidspunkt (side 295 i Falk, 1986). Vi kan si at vi har hendinger med omvendt tidsakse når betingelsen kommer sist i tid.

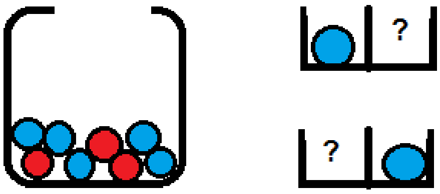
Figur: Viser tidspunkt for betingelse og effekt.
*Teksten **under** pilen viser til at betingelsen kan komme etter effekten i tid.*

Å ta med i diskusjoner og beregninger betingelser som oppstår på «gale» tidspunkt er ikke et valg, men noe som vi skal gjøre (side 292 i Falk 1986).

Feilslutningen om inversjon av tidsaksen er tolkningen av at utfallet av en hending kommer etter den betingede hendingen (årsak – effekt) i situasjoner der det er omvendt (invers).

Ulike forskere kaller slike feilslutninger som «å tolke betingelse som årsak» (side 292 i Falk 1986, med flere). Feilslutningene ble av Falk også referert til som «feiltolkning av tidsaksen» (Diaz og Batanero, 2009) og har også blitt gitt betegnelsen «The Falk Phenomenon» (side 290 i Savard, 2014). Forskerne Gras og Totohasina (1995) identifiserte ulike misoppfatninger om betinget sannsynlighet i en undersøkelse blant 17-18 år gamle elever (Diaz og Fuente, 2007). De kalte misforståelsen med tidsaksen for «the chronological conception», eller misoppfatningen om kronologisk rekkefølge (side 252 i Batanero og Sanchez, 2005).

Eksempel
I en urne er det tre røde og fem blå kuler. Vi trekker tilfeldig to kuler etter hverandre uten å legge kulene tilbake. Oppgavene er:



a) Hvor stor er sannsynligheten for at den andre kulen er blå, når den første kulen var blå?
b) Hvor stor er sannsynligheten for at den første kulen er blå, når den siste kulen er blå?

Løsning:
a) Vifølger tidsaksen og får $P\left(B_2 \middle| B_1\right) = 4/7$.
b) Kort drøfting nedenfor.

Majoriteten av de som svarer på oppgave b gir følgende svar: $P\left(B_1 \middle| B_2\right) = 5/8$. Dette er uriktig tenkning, elever sidestiller dette med $P\left(B_1\right) = 5/8$. Informasjonen fra den andre hendingen reduserer utfallsrommet for den første hendingen og fjerner en blå ball fra de mulige utfall i den første trekningen. Vi får altså: $P\left(B_1 \middle| B_2\right) = 4/7$. Det første spørsmålet i eksempelet ovenfor er kompatibelt med tidsaksen, mens det andre spørsmålet ikke er det. Informasjonen som er gitt i de to oppgavene er i praksis identiske, men med omvendt rekkefølge. Hvem påvirker hvem - oppgavene er symmetriske (side 293 i Falk, 1986). Symmetrien kan dessverre bli forstyrret av semiotikk - hvordan ord skaper mening. I eksempler som ovenfor pekes det på at det er en semiotisk konflikt i bruken av ordene avhengig/uavhengig relatert til tidsaksen (side 123 i Batanero og Diaz, 2007). Eksemplet nedenfor viser mor og datter som begge har blå øyne. Det er mulig å spørre om sannsynlighet for blå øyne hos moren når det er gitt at datteren har blå øyne (side 293 i Falk, 1986).

Eksempel
Selv om andelen blå øyne i to generasjoner, i dette tilfellet mor og datter, er konstante, ble resultatet i en undersøkelse følgende:
En overvekt av de spurte mente at det var større sjanse for at datteren hadde blå øyne dersom moren hadde blå øyne, enn omvendt.

Misoppfatningen konstruerer et hinder for å kunne se den reversible egenskapen ved to hendinger. Oppfatningen om reversibilitet er en nødvendig forutsetning for at elever skal kunne forstå Bayes' setning (side 253 i Batanero og Sanchez, 2005).

2.3.3 3C Bayesiske problemer

Et bayesisk problem er en oppgave som involverer bruk av Bayes' setning, som eksemplet om Downs syndrom og positivt utslag på test i avsnitt 2.3.1. Setningen er gitt ved:

$$P(A|B)=\frac{P(A)\cdot P(B|A)}{P(B)}$$

For å beskrive dette nærmere kan vi se på følgende eksempel:

Eksempel
Nøyaktigheten til en graviditetstest (positivt utslag ved graviditet) kan være så lav som 75 %, når testen er tatt av uerfarne personer (som den ofte er hjemme). Dette betyr at av 100 kvinner som er gravide, vil kun 75 få et positivt testresultat, mens 25 ville fått et negativt resultat. Samtidig er ikke slike tester like nøyaktig den motsatte veien heller. En kvinne som ikke er gravid, kan få et positivt resultat på en slik test.

Dette er et bayesisk problem fordi teori innen sannsynlighet, og spesielt Bayes' setning, er grunnleggende for å undersøke hvordan mennesker anvender og reviderer sin tenkning og tro omkring bestemte situasjoner i livet, som for eksempel graviditet. Et annet eksempel hentet fra New England Journal viser en undersøkelse med en diagnostisk test i et medisinsk spørsmål.

Eksempel
Dersom en test for å undersøke en bestemt sykdom, som forekommer hos 1 av 1000 personer, har en falsk positiv grad på 5 %. Hvor stor er da sjansen for at en person som med positivt utslag har denne bestemte sykdommen? Vi antar ellers likhet mellom alle ukjente faktorer og lignende.
Løsning:
Vi antar at vi har 1000 personer som tar denne testen. En person har sykdommen og vil få positivt resultat på testen. I tillegg vil 50 personer (5 % av 1000) få et positivt resultat, uten at de har sykdommen. Oddsen for at en person med positivt utslag da har sykdommen er 1 : 50 og sannsynligheten er 1/51.

I denne undersøkelsen, som først ble foretatt for om lag 40 år siden, svarte de fleste at det var 95 % sjanse for at personen hadde sykdommen. De som ble spurt var doktorer eller doktorstudenter ved Harvard Teaching Hospital. Når det samme spørsmålet ble gjentatt i 2014, nå på et sykehus i Boston-området, ble ikke resultatet bedre (Martyn, 2014).

2.4 Misoppfatning 4 Synkrone og asynkrone hendinger

Et annet problem knyttet til tidsaksen - misoppfatning er å ta hardt i, er synkrone og asynkrone hendinger. Overskriften kan være noe misvisende siden jeg ikke er direkte på jakt etter forskjellen mellom synkrone og asynkrone hendinger. Jeg søker resultat i prestasjonen til elever før og etter undervisning, målt gjennom oppgaver med synkrone og asynkrone hendinger.

- To synkrone hendinger - skjer samtidig
- To asynkrone hendinger - skjer ikke samtidig

I mange tilfeller er de to situasjonene ovenfor identiske, men teksten i en oppgave kan være formulert som om dette ikke er tilfelle.

Eksempel

a) I urne A er det to svarte og en hvit kule, mens det i urne B er en svart og fire hvite kuler. Vi trekker samtidig ut en kule fra hver urne. Hvor stor er sannsynligheten for at begge er svarte?

b) I en urne er det to svarte og en hvit kule. Vi trekker en kule fra urnen, ser på fargen og legger den tilbake. Så trekker vi en gang til. Hvor stor sannsynlighet er det for å trekke to svarte kuler?

Løsning:

Formelt er de to situasjonene like, selv om tallene er ulike. De to påfølgende forsøkene i b) kan sees på som en samtidig trekning fra to urner, A og B. Hvis det er a svarte og b hvite kuler i urne A og c svarte og d hvite kuler i urne B, er den riktige løsningen gitt ved produktregelen:

$$P(\text{to svarte}) = \frac{a}{a+b} \cdot \frac{c}{c+d}$$

En undersøkelse på 196 elever i alderen 17-18 år viste at mange elever ikke oppfattet denne likheten. I en synkron situasjon var det en tendens til å summere sannsynligheter i stedet for å bruke produktsetningen, for så å bruke den korrekte regelen i en asynkron situasjon (Diaz og Batanero, 2009). De som hadde fått undervisning i bruk av valgtre brukte produktsetningen og presterte bedre i situasjoner med to asynkrone hendinger (Sánchez og Hernandes (2003) i Diaz og Fuente, 2007).

Masteroppgaven vil undersøke prestasjonene i beregninger av $P(B|A)$ (betinget sannsynlighet) og $P(A \cap B)$ (produktsetningen) i situasjoner med og uten tilbakelegg. Graden av prestasjon er knyttet opp mot i hvilken grad elevene forveksler sannsynlighetene ovenfor eller har andre misoppfatninger relatert til synkrone og asynkrone hendinger.

2.5 Misoppfatning 5 Heuristikker

En heuristikk er i denne sammenheng en vurdering som er basert på erfaring. Heuristikken kan være riktig eller feil, avhengig av situasjonen. Ofte, når en heuristikk er nevnt, ligger det implisitt at det er feil anvendelse i den spesielle situasjonen. En og samme person kan ha et vell av heuristikker og kan bruke ulike tilnærminger til noenlunde like problemer.

Hva man observerer sist (eller først) kan ha innvirkning på de valg man tar. Heuristisk representativitet vil si å vurdere sannsynligheten for en hending ut fra hva som tidligere har hendt. Effekten av siste observasjoner, «recency effects», deles inn i to retninger:

- Positiv – forventer samme utfall etter en lang serie (trenden holder seg)
- Negativ – forventer motsatt utfall etter en lang serie (trenden snur)

Gambler's Fallacy er troen på at sannsynligheten for en hending synker og at trenden snur, dersom hendingen har hendt ofte i de senere forsøkene, selv om man ved neste korsvei vet at alle hendingene er uavhengige (side 292 i Savard, 2014). Eksemplet nedenfor viser to retninger for «recency effects» (Madsen, 1995).

Eksempel

Du har kastet en 5-krone 6 ganger og får følgende rekkefølge: MMMMMM.
Hva kan du forvente neste gang du kaster 5-kronen? (oppgave 3 i Madsens test)

		Svarprosent
A	K	0,250
B	M	0,345
C	Blant a-b har alternativene like stor sannsynlighet	0,340
D	Usikker	0,065

Vi ser at tilbøyeligheten til å tro at vi er inne i en god trend, er sterkere enn troen på at nå må trenden snu (Gambler's fallacy). Dog er ikke forskjellen stor. Likevel må vi konstatere at det er bare om lag en tredjedel som gjenkjenner en situasjon med uavhengige hendinger.

En person som anvender representasjonsheuristikk vil lage en rekkefølge med mynt/krone i en serie med myntkast med flere alterneringer mellom mynt og krone, enn det som teoretisk ville skjedd i et tilfeldig forsøk. På samme måte vil ikke personen godta en lang rekke med samme hending i samme forsøk. Dette står i motsetning til hva som teoretisk kan hende i en prosess som er tilfeldig (Batanero, et al, 2014). Eksempelet nedenfor viser forskjellen mellom to diagnostiske oppgaver, den ene rettet mot Gambler's fallacy og den andre mot representasjonsheuristikk. De har begge et ønske om balanse, men på ulike måter. Å ønske balanse i problemet nedenfor vil for en person som bruker representasjonsheuristikk i sin tenkning være styrt av at mengden med tre av hvert kjønn er den rekkefølgen som best kjennetegner hele populasjonen, utvalget representerer populasjonen på den beste måten (side 145 i Konold, 1991).

Eksempel

La oss se på hendingene der G er gutt og J er jente og denne rekkefølgen av seks fødsler:

a) Vi vurderer rekkefølgen GGGGGG og spørsmålet hvilket kjønn har størst sannsynlighet ved neste fødsel – til en person som anvender Gamblers fallacy i sin tenkning.
b) Vi vurderer rekkefølgen GGGGGG og rekkefølgen GJJGGJ og spørsmålet om hvilken rekkefølge som har størst sannsynlighet – til en person som anvender representasjonsheuristikk.

Vurdering:

a) En person som heller mot Gambler's fallacy ønsker å oppnå balanse mellom kjønnene i det lange løp, etter kjent og egen erfaring. Ved presentasjon av rekkefølgen GGGGGG, ville personen svart at kjønnet til neste barn vil være jente (J).
b) I et valg mellom de to rekkefølgene vil en person som bruker representasjonsheuristikk i sin vurdering søke balanse. Ut fra kjent og egen erfart balanse mellom kjønnene ville denne personen svare at rekkefølgen GJJGGJ er mest sannsynlig.

Selv for korte rekkefølger som den beskrevet ovenfor, vil resonnering etter representasjonsheuristikk søke balanse (side 47 i Tversky og Kahneman, 1974).

Ønsket om balanse mellom utvalg og populasjon kan resultere i representasjonsheuristikker.

Etter min mening vil en person som bruker representasjonsheuristikk i valget mellom GJGJGJ og GJJGGJ, altså tre av hvert kjønn i begge tilfeller, valgt GJJGGJ, fordi den beskriver en tilfeldig rekkefølge bedre enn den «symmetriske» rekkefølgen GJGJGJ. Dette blir også bekreftet av Madsens test (Madsen, 1995).

2.6 Misoppfatning 6 Den mest vanlige misoppfatningen

Som et tillegg til misoppfatningene som allerede er nevnt, kan vi trekke inn «Den mest vanlige misoppfatningen» (Devlin, 2014). Devlin sier at han ikke har telt alle misoppfatningene verden over eller har plukket ut en som han mener forekommer oftest, men han vil gjerne komme med en kandidat, her 6A. Jeg er enig med Drevlin og hans kandidat, men jeg har i tillegg trukket inn en annen kandidat, 6B, som jeg mener fortjener litt omtale, og min lanserer min egen til slutt: 6C. Misoppfatningene i kapittel 2.6 er heuristikker.

2.6.1 6A Bestemt sannsynlighet, anchoring I

Den største misoppfatningen innen sannsynlighetsregning er at mange tror at det til en hver begivenhet eller hending, som skjer hvor som helst i verden, følger en bestemt sannsynlighet. Dette har sitt opphav i de mange oppgavene som er laget med terningkast, myntkast, trekke kort fra en kortstokk og så videre. Alle disse leder til troen om at sannsynlighet er en empirisk egenskap i forsøket, heller enn vår viten om utfallene (Devlin, 2014). På engelsk går denne misoppfatningen under navnet «anchoring» og kjennetegnes ved at personer holder fast ved en sannsynlighet, eller en bearbeidelse av denne. Å holde fast ved en metode er en heuristikk, og når metoden ikke virker har vi en heuristisk skjevhet. Sannsynligheten som er forankret kan komme fra tidligere kunnskap eller kan være fra en gitt oppgave (side 289 i Savard, 2014). Eksemplet nedenfor viser problemet med troen på en fast verdi for sannsynlighet i et eksperiment.

Eksempel
Jeg har to barn og et av dem er ei jente.
Hvor stor er sannsynligheten for at mitt andre barn er en jente?

Løsning:
En person som støtter seg til anchoring vil svare ½. Det riktige svaret er 1/3.

Jeg kan selv tenke meg følgende scenario:

Eksempel
Jeg triller en terning foran to personer, den ene med bind for øynene, den andre ikke. Terningen er trillet og jeg spør begge: «Hvor stor er sannsynligheten for at terningen viser 6 øyne?»

Løsning:
Dersom begge forsøkspersonene kan noe om sannsynlighet så skal personen med bind for øynene svare 1/6, mens personen uten bind for øynene skal svare 0 eller 1. Enten ble det en sekser, eller så ble det ikke.

Lærere og andre bedriver en undervisning basert på frekvenser. Faren ved dette er at vi får et fast tall knyttet til sannsynlighet og at denne ikke vil endre seg ved ny kunnskap (Devlin, 2014). Den bestemte sannsynligheten knyttet til et enkelt forsøk blir gjerne justert i et sammensatt forsøk, men i bunn ligger troen på en fast sannsynlighet. Det som gjør sannsynlighetsregning nyttig, er å kunne endre sannsynligheter etter hvert som ny kunnskap oppstår. Anchoring er et hinder for denne nytten.

2.6.2 6B Lik sannsynlighet

Dette blir på engelsk kalt «equiprobability bias» og beskrives ved at personer betrakter to utfall som like sannsynlige når de faktisk ikke er det (side 121 i Batanero og Diaz, 2007). Flere forskere som Lecoutre, Cordier og Durand (1988) og Lecoutre (1992) har studert tendensen til å tenke på hendinger i tilfeldige forsøk til å være av lik sannsynlighet. Det vil si at personer fordeler sannsynligheten jevnt over utfallsrommet, selv om så ikke er tilfelle (side 248 i Batanero og Sanchez, 2005). Et eksempel med sannsynligheten for to eksperimenter (side 560 i Lecoutre, 1992):

Eksempel
To terninger blir kastet samtidig. Hvilken av hendingene har størst sannsynlighet?

a) En femmer og en sekser
b) To seksere
c) De to hendingene ovenfor er like sannsynlige

Løsning: Alternativ a har dobbel så stor sannsynlighet som alternativ b

---

Mange av deltakerne svarte at sannsynligheten var den samme, og at det rette svaret var c. De har en intuisjon om sannsynligheten 1/6 for terningkast generelt. Lecoutre kalte denne feilslutningen for «equiprobabilty bias» og sa at denne feilslutningen har stor motstandskraft mot undervisning og andre forklaringer. Andelen elever som har gitt sine svar basert på lik sannsynlighet er stabil over alle alderssegmenter og i ulike settinger av problemet (side 532 i Fishbein et al, 1991). Ulike forsøk på å rive ned denne robuste misoppfatningen er blitt prøvd, blant annet ved å fargelegge terningene, men uten større hell (Madsen, 1995). Eksempel fra Madsens test:

Eksempel

En vanlig rettferdig terning er malt slik at en side er svart og de andre sidene er gull. Dersom terningen blir kastet en gang, hvilken farge vil du gjette kommer opp?

	Alternativ	Svarprosent
a	Sort	0,070
b	Gull	0,470
c	Det er ikke tilstrekkelig informasjon	0,175
d	Blant a-b har alle alternativene like stor sannsynlighet	0,230
e	Usikker	0,055

Vi ser at andelen elever som baserte sine svar på lik sannsynlighet tilsvarer 23 %.

Det er noen oppgaver i testen der feil kan bli tolket i lys av misoppfatningen om lik sannsynlighet, men det blir drøftet i kapittel 5.

2.6.3 Bestemt fremgangsmåte, anchoring II

På samme måte som elevene knytter bestemte sannsynligheter til bestemte forsøk, som terningkast og kort fra kortstokk, overfører elevene også fremgangsmåter fra oppgave til oppgave. Vi kan si at elevene generaliserer en bestemt fremgangsmåte til å gjelde flere steder. For eksempel erfarer jeg at elever ofte bruker produktsetningen i problemer med to hendinger, enten setningen skal brukes eller ei. Jeg lanserer denne som min egen misoppfatning. Ikke for at jeg skal personifisere en misoppfatning, men for å trekke frem en misoppfatning jeg kun har funnet delvis beskrevet i litteraturen. Med dette mener jeg at jeg tenkte på dette før jeg fant det i litteraturen, og at mengden jeg har funnet ikke er overveldende. Men jeg fant da noe:

> *«Lærere erfarer at en de mest vanlige antagelsene elever gjør, er at sannsynligheten for en sammensatt hending kan beregnes ved å multiplisere sannsynlighetene for hver enkelt hending.»* (side 5 i Tomlinson og Quinn, 1997. Min oversettelse).

De generaliserer en ny kunnskap inn i situasjoner der denne ikke skal brukes. Et eksempel er å bruke produktsetningen der problemet er å finne en betinget sannsynlighet. Dette er notert mer generelt under misoppfatning 1 Kontekst. Hvorvidt vi støter på denne misoppfatningen, eller heuristikken, som jeg ville sagt, senere i denne masteroppgaven skal være usagt, men jeg holder øynene åpne. Det kan også være at jeg lar denne være under misoppfatning 1 og ikke studerer den spesielt. Jeg erfarer også at elevene i kombinatoriske problemer bruker innarbeidede prosedyrer og regneregler som nødvendigvis ikke kan overføres til alle oppgaver.

2.7 Misoppfatning 7 Kombinatoriske problem

Kombinatoriske problem kan lede til mange ulike feil. Her er en liste med kombinatoriske feil (Batanero med flere (1997) side 244 i Batanero og Sanchez, 2005), og rimelig fritt oversatt av meg:

- Rekkefølge – klarer ikke å avgjøre når det er et ordnet eller et uordnet utvalg.
- Tilbakelegging – vet ikke når det er med eller uten tilbakelegging.
- Objektfeil – klarer ikke å skille mellom identiske objekter, eller motsatt.
- Ekskludering – tar vekk elementer i et utvalg.
- Ufullstendig utfallsrom – prøver og feiler med mange utfall, men får ikke med alle.
- Feil formel – bruker feil formel på et problem som er riktig observert.
- Mening med formel – husker ikke hva som er hva, for eksempel n og r.
- Feil ved valgtre – bruker valgtre feil, eller er ikke i stand til å lage et riktig valgtre.

Oppgave 8 i testen er delvis et kombinatorisk problem. Jeg har sett på denne oppgaven i testen slik at det, selv om det er et kombinatorisk problem, i større grad er kunnskapen om å konstruere det riktige utfallsrommet.

2.8 Andre misoppfatninger

Det er andre heuristikker som blir omtalt senere i masteroppgaven, men som ikke står like sentralt. Jeg skriver kort om disse misoppfatningene her. Dette er gjort for å lette drøftingen i kapittel 5 og oppsummeringen i kapittel 6.

2.8.1 8A Tilgjengelighet

Tre vurderingsskjevheter som representasjonsheuristikken ofte fører til er ignorering av utvalgsstørrelsen, ignorering av baserate (se 2.8.3) og konjunksjonsfeilen (se 2.2). Å ignorere utvalgsstørrelsen kan for eksempel være å estimere sannsynligheten for en hending etter hvor sterk troen er på at selv små utvalg representerer hele populasjonen (side 247 i Batanero og Sanchez, 2015). Alle er syke fordi ens nærmeste er syke. Sannsynligheten for en hending er basert på hvor lett det er for en elev å komme på bestemte tilfeller av hendingen. Hvor lett noe kommer frem som løsning i sin egen tenkning, kalles tilgjengelighetsheuristikk (side 151 i Konold, 1991).

Eksempel

Å tro at det er lett å vinne i Lotto, kan falle lettere for en som er fra Verdal, enn for en person fra en annen kant av landet. Grunnen til dette er at det er stor sannsynlighet for at folk fra Verdal kjenner en lottomillionær, og altså har en nærhet til hendingen, slik at den lettere dukker opp fra minnet.

Eksempel

Å være vitne til en ulykke kan påvirke graden av tro på at en selv kan bli utsatt for en ulykke.

Å se hva man ikke forventer å se, kan i større grad påvirke bedømmingen av sannsynligheten av det man så, altså en overestimering av sannsynligheten til en hending. Det man gjør først og sist husker man best, altså en forbindelse til «recency effects».

2.8.2 8B Utfallstilnærming

En representasjonstilnærming til sannsynlighet studerer et utvalg av forsøk, mens en utfallstilnærming korter dette ned til å se på sannsynligheten i et enkelt forsøk. Denne tenkningen er beskrevet av Konold (1989), som også kaller dette for «outcome approach», eller utfallstilnærming til usikkerhet (side 392 i Konold et al, 1993). I «outcome approach» er hovedmålet i sannsynlighet å forutsi utfallet i et enkelt forsøk. La oss se på et eksempel for å skille representasjonsheuristikk fra utfallstilnærming (side 393 i Konold et al, 1993).

Eksempel

En terning har en hvit side og fem svarte sider. Terningen trilles seks ganger. Hvilket av alternativene nedenfor er mest sannsynlig?

a) Svart side opp på fem kast og hvit side opp på ett kast.
b) Svart side opp på alle seks kastene.

Løsning:
Beregninger viser at sannsynligheten for a er 0,402 og sannsynligheten for b er 0,335.

En person som svarer alternativ a kan ha basert sitt svar på representasjonsheuristikk. Det er basert på en vurdering av hvor godt de to ulike utvalgene representerer hele populasjonen. Forholdet mellom hvite og svarte sider i utvalget, svarer til forholdet mellom svarte og hvite i populasjonen (side 394 i Konold et al, 1993). Med populasjon menes her at ett av seks kast ville i det lange løp gitt hvit side opp. Hva da med de som svarte alternativ b? En teori kan være at de brukte utfallstilnærmingen. De ser ikke at de skal spesifisere en sannsynlighet som reflekterer fordelingen som kan forekomme i et utvalg, men si hva som skjer i et enkelt utfall. I et enkelt utfall er det størst sannsynlighet for svart side opp. Deretter gjentar de dette svaret seks ganger. En slik tenkning er basert på utfallstilnærming, et slags ja-eller-nei svar i hvert enkelt forsøk (side 395 i Konold et al, 1993).

2.8.3 8C Base Rate

Angående tolkning som trengs for å løse problemer som involverer Bayes' setning, viser forskning at elever ikke anvender Bayes' setning intuitivt. Forskningen viste at «base rate fallacy» stod godt og stødig. Base rate fallacy er feilslutningen som oppstår når den som skal utføre beregningene neglisjerer størrelsen på populasjonen, den underliggende informasjon, og bare studerer sannsynlighetene. Denne feilslutningen ble identifisert når elever ble bedt om å løse et problem som involverte Bayes' setning (Diaz og Fuente, 2007). Å studere størrelsen på populasjonen er altså viktig. Eksempelet på neste side viser at det kanskje ikke er så rart at man tar feil likevel, og at den underliggende informasjon må studeres nøye. Dette er kjent som Simpsons paradoks (side 88 i Lillestøl, 1991).

Eksempel

Tabellen til venstre nedenfor viser en studie over behandling av nyrestein, store som små, der behandling B fremstår som det beste alternativet totalt sett.

Behandling A	Behandling B
Små steiner < 2 cm - 93 %	Små steiner < 2 cm - 87 %
Store steiner > 2 cm - 73 %	Store steiner > 2 cm - 69 %
Begge - 78 %	Begge - 83 %

Behandling A	Behandling B
Små steiner: 81/87 -> 93 %	Små steiner: 234/270 -> 87 %
Store steiner: 192/263 -> 73 %	Store steiner: 55/80 -> 69 %
Begge: 273/350 -> 78 %	Begge: 289/350 -> 83 %

Tabellen viser at det samlede resultatet er best med behandlingsmåte B, mens behandlingsmåte A har best resultat for både store og små steiner. Åpenbart, eller nesten åpenbart, må A være den beste behandlingen, men hvordan kan det ha seg at B har høyest score når vi ser på det samlede resultat? Svaret ligger i populasjonen til hver enkelt behandling som vi ser i tabellen til høyre.

Fra helsevesen i eksemplet ovenfor til rettsvesen i eksemplet nedenfor. Eksemplet setter fokus på relevansen til informasjonen om «base rate» (side 63 i Watson og Moritz, 2002)

Eksempel

En taxi var involvert i en kriminell handling i en by. To taxi-selskaper opererer i denne byen, Grønn og Blå, og et vitne identifiserte taxien som Blå. Du får vite følgende fakta:

- Av taxiene tilhører 85 % Grønn og 15 % tilhører Blå.
- Et vitneprov under slike omstendigheter er riktig i 80 % av tilfellene

Oppgaven er å finne sannsynligheten for at taxien som var involvert tilhørte selskapet Blå.

Løsning:

Påliteligheten til vitneutsagnet virker stor, men på grunn av den store andelen grønne taxier (base rate) finner vi, ved hjelp av Bayes' setning, at det riktige svaret er 41 %.

Eksemplet ovenfor er brukt i mange sammenhenger der Bayes setning er sentral. I en undersøkelse med dette eksemplet svarte de fleste at sannsynligheten var over 50 %. Dersom dommere og jurymedlemmer brukte samme tankegang, vil det være store sjanser for at feil person havnet i fengsel (side 7 i Tomlinson og Quinn, 1997). Jeg er usikker på i hvilken grad vi kommer inn på denne misoppfatning senere , men det blir drøftet i kapittel 5.

3 Metode

Ordet metode stammer fra gresk; methodos – bestemt vei mot mål. Kapitlet inneholder generell teori om metode og metoden som er brukt i undersøkelsene til Diaz, Batanero og Fuente sin CPR test. Jeg begrunner mine valg av metoder og vurderer kvaliteten på undersøkelsen.

3.1 Teori om metode

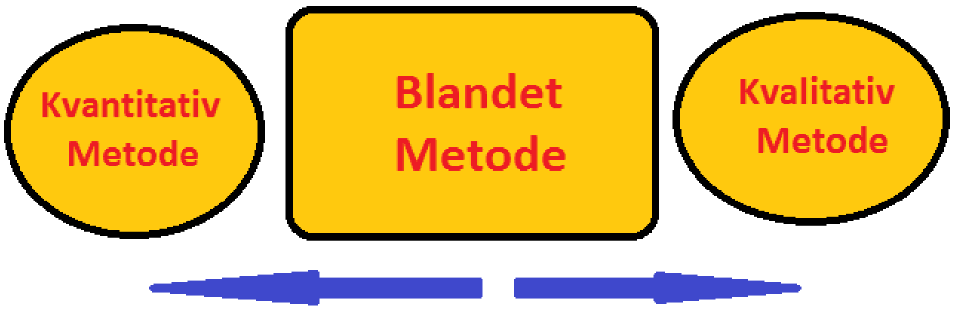
En metode er en fremgangsmåte basert på regler og prinsipper og brukes særlig i vitenskap og forskning. I lang tid har det hersket to hovedtyper av metode; *kvantitativ* og *kvalitativ* metode. Forskningen har vært preget av at man enten brukte den ene eller den andre metoden:

> *«It also includes designing and writing the research in one of the two major tracks: quantitative research or qualitative research.»* (side 11 i Creswell, 2012)

I de senere tiår har en blanding av de to metodene blitt benyttet i større grad, og dermed fått sitt eget navn. Metoden kalles *mixed-methods*, eller blandet metode på norsk.

> *«A research paradigm whose time has come.»* (side 14 i Johnson og Onwuegbuzie, 2004)

Ordet «paradigme» beskriver noe stort, mønstergyldig og eksempel til etterfølgelse. Mange forskere taler varmt for blandet metode og mener at det var på tide at metoden fikk større plass. Dersom vi plasserer kvantitativ metode og kvalitativ metode i hver sin ende, kan vi si at blandet metode, som dekker området mellom de to ytterpunktene, har stor plass.



Figur: Kvantitativ, kvalitativ og blandet metode.

En undersøkelse viste at utbredelsen av mixed-method går sakte, men sikkert fremover. Mens ren kvalitativ metode sank fra 60 til 45 prosent, økte mixed method fra 10 til 20 prosent fra 2005 til 2010 (side 335 i Gellert med flere, 2013).

3.1.1 Kvantitativ metode

Metoden baserer seg på innsamling og analyse av til dels store mengder data som både må og skal tallbehandles. Resultat fra undersøkelser basert på kvantitativ metode presenteres ofte i tabeller, grafer og diagrammer. Regneark program er særdeles egnet under innsamling, analyse og presentasjon av en undersøkelse med kvantitativt design. Bildet nedenfor viser et utsnitt av en av de mange sidene jeg hadde i programmet excel i forbindelse med denne oppgaven:

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
A - S1			**Antall deltakere**					**n=**	**21**								
Oppgave	**1a**	**alt**	**1b**	**alt**	**1c**	**alt**	**1d**	**alt**		**2**			**3**			**4**	
Kandidat	1	svar	1	svar	1	svar	1	svar	i	ii	iii	a	b	c	a	b	c
1	1			75/350	1		1				1			1			1
2	1			2,90 %	1			75/780		1				1			1
3	1		1		1		1			1				1		1	
4	1			75/104	1			104/75		1				1	1		
5	1		1		1		1			1				1			1
6	1		1		1		1			1			ubes				1
7	1		1		1		1			1				1			1
8	1		1		1		1			1				1			1
9	1			75/104		35/78		75/350			1	1				1	
10	1		1		1		1			1			1				1

Figur. Analyse av data fra en undersøkelse i forbindelse med oppgaven.

Nedenfor følger tre definisjoner av kvantitativ metode som jeg har funnet på internett:

> Kvantitative metoder er forskningsmetoder som befatter seg med tall og det som er målbart (kvantifiserbart). (kilde: https://no.wikipedia.org/wiki/Kvantitativ_metode)

> Quantitative research is a formal, objective, systematic process in which numerical data are used to obtain information about the world.
> (kilde: http://www.researchproposalsforhealthprofessionals.com/definition_of_quantitative_resea.htm)

> The use of sampling techniques (such as consumer surveys) whose findings may be expressed numerically, and are amenable to mathematical manipulation enabling the researcher to estimate future events or quantities. (kilde: http://www.businessdictionary.com/definition/quantitative-research.html)

Statistikk vil si å bruke et utvalg til å si noe om en hel populasjon. Dette harmonerer også med de to siste definisjonene av kvantitativ metode. Kvantitativ metode tar ofte utgangspunkt i hypoteser og vil enten avkrefte eller bekrefte disse. Kvantitative metoder brukes til å analysere trender, sammenligne grupper med relaterte variabler i en statistisk analyse (side 13 i Creswell, 2012). Tilnærmingen er objektiv og rett frem. Ofte brukes et spørreskjema med faste spørsmål og tilhørende valgmuligheter (multiple choice). Dersom spørsmålene er mer åpne, brukes det ofte kategorier for de ulike svarene objektene kan gi. Det må da herske enighet om hvilke svar som er delvis riktig og hvilke som er feil, og gjerne også hvilke svar som kan godkjennes som riktige, selv om de ikke er helt perfekte.

Styrker	Svakheter
Kan teste og bekrefte gitte teorier for hvordan fenomener oppstår	Enkelte fenomener kan glippe fordi undersøkelsen ikke tar høyde for dette
Kan teste fremsatte hypoteser	Vanskelig med lokal tilpassing
Kan generalisere med bakgrunn i presise numeriske data	Generalisering kan hindre direkte bruk av resultatet på for eksempel enkeltpersoner
Rask dataanalyse med dagens programvare	Trenger stor datamengde (egen)
Forskningsresultatene er uavhengige og ofte lett reproduserbare	
Kan ha høy kredibilitet ved mange deltakere	

Tabell. Styrker og svakheter med kvantitativ metode (side 19 i Johnson og Onwuegbuzie, 2004).

Kvantitativ metode er underlagt formelle krav som danner grunnlag for en stram prosess fra begynnelse til slutt. Kjente talere for kvantitative metoder, såkalte kvantitative puritanere er Ayer, Maxwell og Delaney, Popper og Shrag (side 14 i Johnson og Onwuegbuzie, 2004).

3.1.2 Kvalitativ metode

Mens kvantitativ metode kjennetegnes ved spørreskjema og data fra mange personer, kjennetegnes kvalitativ metode ved intervju eller observasjoner av enkeltpersoner eller mindre grupper. Siden antallet forsøkspersoner med kvalitativ metode er betydelig mindre enn ved kvantitativ metode, er vi også i større grad avhengig av personene som skal intervjues eller observeres er nøye utvalgt. Det må være et *hensiktsmessig utvalg*, som skal kunne gi svar på fenomener og forskningsspørsmål. På den motsatte siden av skalaen finner vi *tilfeldig utvalg*, som i større grad brukes i kvantitativ metode. Med tilfeldig utvalg, menes det ikke at vi spør folk på Mars hvordan livet er på jorden, men mer at «alle som er i gruppen» får være med på undersøkelsen, og at gruppen er representativ. Jamfør diskusjonen rundt elever som blir holdt borte når nasjonale prøver skal gjennomføres i skolen.

Tilfeldig «kvantitativ» utvalg	Hensiktsmessig «kvalitativt» utvalg
Representative objekter i utvalget. Utvalget representerer en større populasjon.	Objektene i utvalget skal i størst mulig grad hjelpe oss til å beskrive fenomener. Skape en detaljert forståelse av fenomenet.
Si noe om populasjonen basert på utvalgets respons. Bygge teorier.	Kan gi nyttig informasjon Kan hjelpe til å forstå et fenomen Kan gi stemme til andre (som lider av samme fenomen, for eks dyskalkuli)

Figur. Hensiktsmessig og tilfeldig utvalg. (Figur side 206, Creswell, 2012)

I kvalitativ forskning er det typisk å studere få individer eller grupper for å kunne gå i dybden hos den enkelte. Dess flere som er med, jo mer blir dybden og kompleksiteten til den enkelte visket ut. Nedenfor følger tre definisjoner av kvalitativ metode som jeg har funnet på internett. For de to første definisjonene har jeg brukt samme nettsted som for de to første definisjonene til kvantitativ metode.

Kvalitativ metode er en metode for generering av kunnskap hvor man undersøker hvilken mening hendelser og erfaringer har for de som opplever dem, og hvordan de kan fortolkes eller forstås også av andre. (Kilde: https://no.wikipedia.org/wiki/Kvalitativ_metode)

Qualitative research consists of a number of differently developed methods that are best suited to address questions of particular interest. (Kilde: http://www.researchproposalsforhealthprofessionals.com/definition_of_qualitative_resear.htm)

research that is not simply the collecting of statistics, but which focuses on reasoning and cultural and social factors, which are researched and then analysed. (Kilde: http://www.dictionarycentral.com/definition/qualitative-research.html

Det finnes flere kjente talsmenn for kvalitative metoder, som for eksempel: Campbell og Stanley og Lincoln og Guba (side 14 i Johnson og Onwuegbuzie, 2004).

Styrker	**Svakheter**
Data er tilpasset den lokale situasjonen og deltagers nivå og kunnskap	Resultatet i en undersøkelse kan være lite nyttig i andre situasjoner
Dypstudie av færre	Vanskelig å si noe bestemt om en større gruppe
Beskrive komplekse fenomener	Vanskelig å teste hypoteser og teorier
Gir forståelse og beskrivelse av personlige erfaringer av fenomener i tilpasset miljø	Det tar generelt mer tid til å samle, bearbeide og analysere data og tolke resultater
Mulig å oppklare misforståelser og korreksjon av data	Resultatene er i større grad influert av forskerens ståsted og behov.
Mulig å studere dynamiske prosesser og gjøre nødvendige endringer underveis	Har lavere kredibilitet hos øvrigheter
Kan i større grad undersøke hvorfor, i stedet for hvilke, fenomener som oppstår.	

Tabell. Styrker og svakheter med kvalititativ metode (side 20 i Johnson og Onwuegbuzie, 2004).

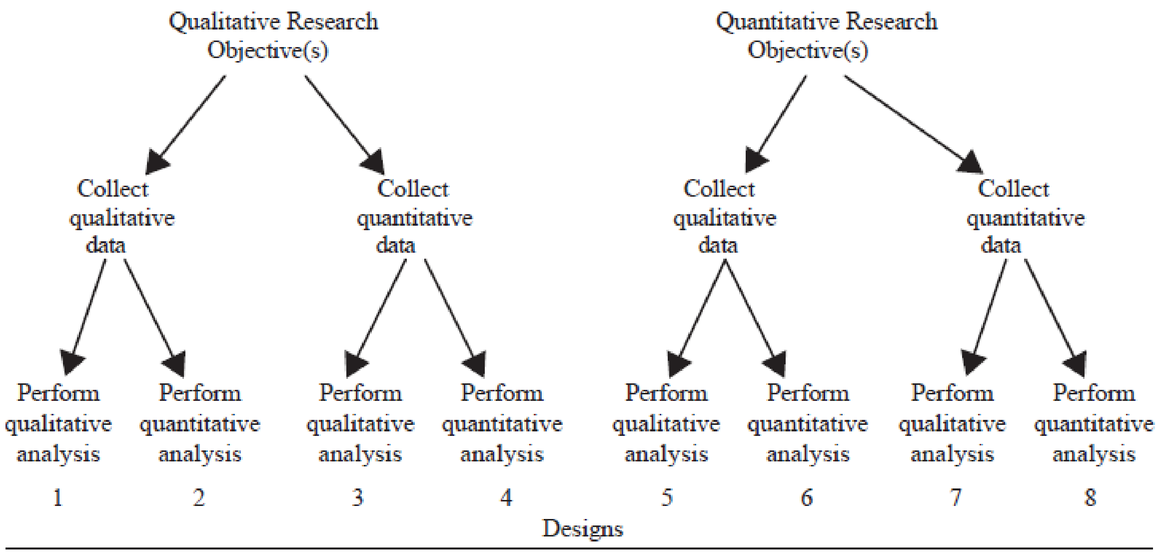
I kvantitativ metode er bruk av observasjon og intervju to hovedretninger til å skaffe informasjon. Observasjoner kan gjøres på flere måter, men grovt sett varierer dette med graden av deltagelse. Det gjøres gjerne et skille for den som undersøker mellom det å være deltager og det å være observatør (side 214 i Creswell, 2012). Et intervju kan gjennomføres på mange ulike måter, med og uten spørreskjema. Et kvalitativt intervju er når en forsker spør en eller flere deltakere åpne spørsmål og tar vare på svarene som blir gitt. Hovedretninger av intervju (side 213 i Creswell, 2012):

- Ustrukturert og åpent intervju med notater
- Ustrukturert og åpent intervju med lydopptak og transkripsjon
- Semistrukturert intervju med lydopptak og transkripsjon
- Fokusgruppe intervju med lydopptak og transkripsjon
- Åpne tilbakemeldinger til et elektronisk intervju eller spørreskjema
- Åpne tilbakemeldinger til oppgaver på et spørreskjema

Spørsmålene skal være åpne slik at stemmen til den enkelte deltaker blir hørt uten påvirkning av for eksempel ledende spørsmål. Påstander og svar i en kvantitativ sammenheng kan få nytt innhold i en kvalitativ sammenheng. Man kan lettere få fatt i tankene til elevene.

3.1.3 Blandet metode

Multippel metode inkluderer mer enn én metode for datainnsamling og forskning. Blandet metode, mixed methods, er en multippel metode som blander kvantitativ og kvalitativ metode.
Figuren nedenfor viser stringent kvalitativ metode helt til venstre og stringent kvantitativ metode helt til høyre. Alt som ligger mellom er mixed-methods (side 21 i Johnson og Onwuegbuzie, 2004).



Note. Designs 1 and 8 on the outer edges are the monomethod designs. The mixed-model designs are Designs 2, 3, 4, 5, 6, and 7.[10]

Figur. Blandet metode design med 1 og 8 som ren kvalitativ og ren kvantitativ metode.

Enhver blanding mellom de to metodene er altså blandet metode. Blandingen kan helle mot den ene eller den andre metoden, ved vektlegging eller tidspunkt. For en mer nyansert beskrivelse av blandet metode (kvantitativ og kvalitativ) må vi se på:

- Graden av hver metode
- Rekkefølgen til metodene
- Analysering av data fra de to metodene
- Miksing av data fra de to metodene

Svarene på disse punktene gir en beskrivelse av designet til metoden. Både kvantitativ og kvalitativ metode har sine styrker og svakheter. Hensikten med å blande dem er ikke å erstatte en av dem eller begge metodene, men heller å øke styrken og minske svakhetene til begge metoder. En forkortet styrke/svakhet versjon for blandet metode (side 21 i Johnson og Onwuegbuzie, 2004):

Styrker	Svakheter
Resultatene kan utdype hverandre	Bruker lang tid
Dypere forståelse av problemstillingen og forskningsspørsmålene	Stor mengde datamateriale og analyse, det blir verken eller.

Tabell. Styrker og svakheter med blandet metode.

Dagens forskning er kompleks og dynamisk og krever at forskere er i stand til å implementere begge metoder og blandingen av disse, for å øke forståelsen, evnen til tolkning og grunnlaget for samarbeid med andre forskere for å oppnå fremdragende forskning (side 15 i Johnson og Onwuegbuzie, 2004). Forskere som fremhever blandet metode er for eksempel: Creswell, Clark, Caracelli og Greene, Tashakkori og Teddlie og Bazeley (side 538 i Creswell, 2012).

3.2 CPR – Conditional Probability Reasoning

I dette kapitlet vil jeg redegjøre for metodene og valgene til Diaz og Fuente (2007) og Batanero og Diaz (2009). Oppgavene som jeg bruker i min undersøkelse er i sin helhet hentet fra en test som disse forskerne har utarbeidet sammen med andre forskere. De har gitt testen navnet Conditional Probability Reasoning (CPR). Testen består av 18 oppgaver innen betinget sannsynlighet. Jeg vil bruke forkortelsen CPR når jeg omtaler testen videre. Når jeg referer til forskerne bak testen, vil jeg noen ganger bruke en egen forkorting: «Diaz med flere». Dette referer til Diaz og Fuente (2007) og Batanero og Diaz (2009). Jeg har laget denne forkortingen blant annet for å bedre flyten i teksten, men vil samtidig vise til referansen i litteraturlisten. Begge artiklene er referert så ofte at sidetallet til hver referanse er utelatt. Grunnen er at artiklene er så sentrale at de bør leses i sin helhet.

Forskningsspørsmålet til Batanero og Diaz (2009) var å evaluere mulige relasjoner mellom formell kunnskap i betinget sannsynlighet og misoppfatninger relatert til dette. Undersøkelsen til Batanero og Diaz er en fortsettelse av forskningen og undersøkelsen til Diaz og Fuente (2007). Batanero og Diaz analyserer mulige forbedringer innen formell kunnskap i betinget sannsynlighet etter instruksjon, altså undervisning i emnet. Som en følge av mulig forbedring ser de etter en reduksjon i misoppfatninger. Til denne analysen brukte de CPR-testen på to grupper med psykologistudenter (n = 177 før undervisning) og (n = 206 etter undervisning). Batanero og Diaz bruker kvantitativ metode.

Til tross for relevansen betinget sannsynlighet har på mange områder og all forskning og litteratur som det finnes om emnet, fant ikke Diaz og Fuente noe velegnet spørreskjema som globalt kunne vurdere studenters forståelse og misoppfatninger innen betinget sannsynlighet. Som en følge av dette konstruerte de CPR-testen, som etter eget utsagn er «et velegnet instrument» for

1. å måle ulike feil og vurdere ulike misoppfatninger relatert til betinget sannsynlighet.
2. å studere innvirkningen undervisning i betinget sannsynlighet har på forekomsten av misoppfatningene.

I tillegg ville de utforske mulige relasjoner mellom oppgavene som måler formell kunnskap og de oppgavene som måler feil som er beskrevet i litteraturen – diagnostiske oppgaver. Batanero og Diaz ville ha et spørreskjema som hindrer ulike tolkninger og forståelser av hva som for eksempel menes med det «å forstå betinget sannsynlighet». For å kartlegge studentenes kunnskaper om betinget sannsynlighet, ble dette vide begrepet dekomponert i flere små enheter. Enhetene ble valgt gjennom analyse av innholdet i 19 ulike lærebøker, plukket ut på bakgrunn av en anbefalt liste fra 31 universiteter i Spania. Alle lærebøker som ble anbefalt av minst fire universiteter ble tatt med i analysen. Etter at bøkene som ikke inneholdt betinget sannsynlighet var fjernet, stod man igjen med 19 bøker. Innholdet ble kategorisert ifølgende enheter:

Nr	Enheter	Oppgave	CPR
1	Definere betinget sannsynlighet og gi et passende eksempel		11
2	Gjenkjenne at en betingelse gir en begrensning av utfallsrommet	8	12
3	Løse bayesisk problem med "Base rate fallacy"	_	2
4	Identifisere betingede, enkle og samensatte sannsynligheter	1	1
5	Identifisere en betinget sannsynlighet og den inverse sannsynligheten	4	7
6	Feilslutningen om snitt i sammensatte hendinger (conjuction fallacy)	3	6
7	Identifisere uavhengige og gjensidig ekskluderende hendinger	_	3
8	Beregne betinget sannsynlighet i ett enkelt forsøk	9	13
9	Beregne betinget sannsynlighet i et forsøk med tilbakelegging	10	15
10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging	2, 6	4, 9
11	Beregne betinget sannsynlighet i et forsøk med union og sammensatte hendinger	_	5
12	Beregne betinget sannsynlighet i et forsøk med omvendt tidsakse	5, 6	8, 9
13	Identifisere betinget årsak- og diagnostisk situasjon	4	7
14	Beregne betinget sannsynlighet i en asynkron situasjon	2, 5, 6, 7	4, 8, 9, 10
15	Beregne betinget sannsynlighet i en synkron situasjon	1, _, _, 12	1, 5, 14, 17
16	Beregne sannsynlighet med produktsetningen på to uavhengige hendinger	11	16
17	Beregne sannsynlighet med produktsetningen på to avhengige hendinger	7, 12	10, 17
18	Beregne total sannsynlighet	_	14
19	Løse bayesisk problem	13	18

Tabell. Nummer og beskrivelse av enheter fra CPR. Oppgavenummer i min undersøkelse og CPR.

Tabellen ovenfor viser for eksempel at oppgave 2, 3, 5, 11 og 14 fra CPR-testen ikke er med i min undersøkelse. Men hva måler oppgavene og enhetene? Dette mener jeg er sentralt i CPR. Inndelingen i enheter gir også begrensninger. For eksempel:

Nr	Enheter	Oppgave	CPR
2	Gjenkjenne at en betingelse gir en begrensning av utfallsrommet	8	12

Denne enheten kan vi finne igjen i mange av oppgavene, men enheten blir bare målt i oppgave 8. CPR-testen måler forståelsen av betinget sannsynlighet, og betingelser gir en begrensning av utfallsrommet. I resultatdelen vil jeg holde meg strengt til enhetene i tabellen gitt ovenfor, men i drøftingsdelen vil jeg trekke tråder mellom enheter og oppgaver som faller utenfor i tabellen ovenfor – for å bekrefte eller avkrefte hypoteser.

Diaz med flere laget en oppgavebank med 49 ulike oppgaver, der noen av oppgaven hadde vært brukt tidligere av andre forskere. Navnene til disse er oppgitt i parentes ved hver oppgave i appendiks. De 49 ulike oppgavene ble så redusert til 18 etter en pilotundersøkelse og i samråd med forskere fra fem ulike land: Brasil, Colombia, Venezuela, Mexico og Spania. Utvelgelsen av oppgavene skjedde etter hvor adekvat hver oppgave var med tanke på (Diaz og Fuente, 2007/Diaz og Batanero, 2009):

a) Hvorvidt oppgaven dekket en eller flere av enhetene i tabellen ovenfor.
b) Om eksperter var enige om at oppgavene var adekvate
c) Oppgavenes vanskelighetsgrad

Oppgavene ble laget i to ulike formater:

1) Multiple-choice-oppgaver med 3-4 svaralternativer (max 1 poeng på hver, oppgave 1 - 10)
2) Åpne oppgaver (max 2 poeng på hver, oppgave 11 - 18)

Etter en ny pilot blant 57 psykologistudenter i 2003-2004, blant annet for å studere CPR sin pålitelighet og validitet, ble det foretatt en ny revisjon av oppgavene. Hver oppgave ble laget i flere eksemplarer med ulik ordlyd. Fra ulike hold ble 13 språkeksperter bedt om å finne den ordlyden i hver oppgave som de syntes var best. Dersom et svar på de åpne oppgavene (11-18) ble ansett som delvis korrekt, ble det gitt ett poeng. Svar som grunnleggende og i sin ide er riktig, men som likevel er feil, er definert som delvis korrekt. CPR-testen ble først gjennomført i en undersøkelse på 414 førsteårs-studenter mellom 18 og 19 år (av Diaz og Fuente) før den på nytt ble gjennomført på 383 studenter (av Batanero og Diaz). Den første undersøkelsen ble brukt som kontrollgruppe for den andre. Det er den siste av disse to undersøkelsene jeg vil vie størst oppmerksomhet i de neste kapitler, men tall fra den første undersøkelsen er tatt med i kapittel 4 Resultat. I den siste undersøkelsen ble studentene delt inn i to grupper, direkte knyttet til forskningsspørsmålet; hjelper undervisning på prestasjon?

- D & B – F tok testen **før** en undervisningsøkt i betinget sannsynlighet, n = 177, (Gruppe 1)
- D & B – E tok testen **etter** den samme undervisningsøkten, n = 206, (Gruppe 2)

Navnene D & B – F og D & B – E brukes for gruppene til Batanero og Diaz, før og etter undervisning. Undervisningsøkten tok for seg betinget sannsynlighet og Bayes’ setning med hjelp av valgtre, krysstabeller og adekvate eksempler. Varigheten av økten står det ingenting om, men den ble avsluttet omtrent to uker før studentene tok testen. For å holde litt orden, må denne undervisningsøkten (som det hele tiden refereres til) ikke forveksles med *introduksjonskurset* **). Begge gruppene fikk et introduksjonskurs i statistikk i begynnelsen av studiet. Bortsett fra at gruppe 1, i motsetning til gruppe 2, ikke hadde hatt undervisning i betinget sannsynlighet, måtte de to gruppene i størst mulig grad være ensartete:

- Første års studenter på psykologi studiet ved Universitetet i Grenada og Murcia
- De fleste studentene var i alderen 18 – 19 år
- Studentene hadde ulik sosial bakgrunn
- Det var 60 prosent jenter og 40 prosent gutter, som er normalt på psykologistudier i Spania
- Alle studentene har hatt undervisning i betinget sannsynlighet i den videregående skole, omtrentlig alder er da 14 – 15 år.
- Alle studentene hadde fulgt det samme introduksjonskurset i statistikk før de tok CPR-testen.

***) Introduksjonskurs i Statistikk: fordelinger, gjennomsnitt, bivarians, korrelasjon og regresjon. Sannsynlighetsregning: enkle eksperimenter, utfallsrom og Laplace-regelen.*

Jeg vil tillegge at noen av oppgavene avslører bestemte misoppfatninger, mens andre oppgaver kan lede til en mengde ulike misoppfatninger. I det siste tilfellet kan jeg gjennom mixed method i noen tilfeller avgjøre hvilken misoppfatning elevene har lent seg mot i sin tolkning. Oppgavene i den kvantitative delen er altså nokså ulike i sitt tolkningsrom. Jeg vil noen ganger bruke «CPR-undersøkelsen» eller bare «CPR», det vil si CPR-testen til Diaz med flere.

3.3 Min metode

Min metode er blandet metode, mixed method. Dette kan for eksempel kjennetegnes ved at fra kvantitativ metode har jeg en test (CPR) med oppgaver for et stort utvalg elever, mens jeg fra kvalitativ metode har intervju med et mindre utvalg av elever. Det totale antallet elever som har tatt testen er 136 og det totale antallet intervjuobjekter er 10.

3.3.1 Valg av metode

Hvorfor mixed method? I første omgang hadde jeg sett for meg en ren kvantitativ metode med stort tallmateriale, men jeg ble samtidig litt skremt av de formelle kravene metoden stiller samt at undersøkelsene til Diaz med flere er kvantitative metoder. Jeg ville gjøre noe annet.

Når vi samler inn data og skal si noe om disse, kan p-verdien oppleves som et orakel som bedømmer våre resultater (1462 i Pripp, 2015). En p-verdi tar utgangspunkt i to hypoteser, der den ene er «Null-hypotesen». Null-hypotesen har som utgangspunkt ingen endring. Relatert til forskningsspørsmålet er null-hypotesen at undervisning ikke gir en forbedring av resultatene i sannsynlighetsregning, selv om resultatene rent tallmessig skulle vise at så var tilfellet.

Vi går da ut fra at dette skyldes tilfeldig variasjon, og at det dermed ikke er en reell statistisk forskjell. I motsetning til null-hypotesen har vi den alternative hypotesen, som da sier at null-hypotesen ikke er sann. Sannsynligheten for det som er målt eller observert, under forutsetning av at null-hypotesen er sann, kaller vi for p-verdien. Dersom p-verdien er lavere enn en på forhånd gitt verdi, som oftest 0,05, sier vi at resultatet er statistisk signifikant – og null-hypotesen forkastes. Å forkaste null-hypotesen i denne oppgaven vil si at vi får en bekreftelse på at undervisning i sannsynlighetsregning øker prestasjonen til elevene. Vi må da selvfølgelig ha en presis beskrivelse av hva denne undervisningen innebærer – for undervisning i seg selv har et utall av variabler. Æren for p-verdien 0,05 tilskrives statistikeren Ronald A. Fisher (1890 – 1962) og er en studie er i seg selv.
Å måle en p-verdi gir ikke et svar på graden av økende prestasjon. Den statistiske definisjonen av p-verdi er ganske kompleks og det er mange fallgruver ved å bruke denne (1462 i Pripp, 2015). I Den Danske Ordbog er p-verdien beskrevet som «ofte tvetydig og kræver tolkning af særlig kyndige», altså enda et varsel om å være forsiktig med å bruke ren kvantitativ metode på et utvalg med få deltakere. I tillegg viser jeg til en artikkel på side 35 i Aftenposten som sier «Stor ståhei for liten p-verdi; Norske forskere er for dårlige i statistikk» (Thoresen, 5. mars 2015). Derfor ikke ren kvantitativ metode, men jeg var heller ikke fristet til å hoppe helt til den andre siden av veien - ren kvalitativ metode. Jeg ville beholde et sammenligningsgrunnlag med tallene til CPR-testen. Jeg ville ha sammenligninger gjort under de samme betingelsene – men jeg ville ha begge metoder. Samtidig som jeg argumenterte for vanskelighetene forbundet med en ren kvantitativ metode ovenfor, satte jeg frem vanskelighetene forbundet med en ren kvantitativ metode. Jeg ville ikke kopiere, men jeg ville heller ikke gjøre noe komplett nytt, når jeg først hadde fått hånd om CPR-testen. Dette er i hovedsak årsaken til at jeg landet på blandet metode. Samtidig som jeg vurderte vanskelighetene med hver metode, synes jeg at jeg fikk svar tilbake i form av fordelene til hver metode. Jeg hadde en tanke om at det skulle være reproduserbart, at det skulle være mulig å få svar på spørsmålene på et annet tidspunkt og et annet sted. Det er større sjanse for dette med kvantitativ metode og spørreskjema. Å utnytte denne styrken til kvantitativ metode, gjorde at jeg kunne gjennomføre en undersøkelse på en annen skole ledet av en annen person.

Listen nedenfor viser andre tester innen sannsynlighet med forsker, tidspunkt og alderssegment:

- Green: 1989 – alder 11 – 16, 1989 – alder 13 – 14, 1991 – alder 13 – 14
- Toohey: 1995 – alder 11 – 15
- Batanero og Serrano: 1999 – alder 14 – 17
- Pratt: 2000 – alder 10 – 11

CPR-testen var rettet mot betinget sannsynlighet, noe som passet med min problemstilling. Nivået på denne testen passet til mine elever, men er elevene et representativt utvalg?

3.3.2 Utvalget

Med en kvantitativ og en kvalitativ del i min undersøkelse måtte jeg ha utvalg av elever som passet til bestemte krav. Utvalg til den kvantitative delen skal være av mer tilfeldig art, men de skal samtidig være representative. Dette utvalget er elever på vg1 og vg2 nivå på den videregående skole – studiespesialiserende, i alderen 15 – 17 år. Jeg har gjennomført syv kvantitative undersøkelser som beskrives i tabellen nedenfor. Kolonnen før/etter viser til om dette er før eller etter undervisning i sannsynlighet, på det aktuelle nivå. Senere i oppgaven vil jeg bruke navnet til de ulike gruppene slik det står i femte kolonne.

Gruppe	Tidspunkt	Antall	Beskrivelse – internt bruk	Navn (videre)	Nivå	før/etter
Pilot	jan 2015	16	Pilot – 1F BKS (14-15) *	**Pilot – F**	vg1 - 1T	Før
1	mars 2015	14	Undersøkelse 1 1G BKS(14-15)	**Und1 – E**	vg1 - 1T	Etter
2	mars 2015	14	Undersøkelse 2 M - S1 BKS (14-15)	**Und2 – E**	vg2 - S1	Etter
3	mars 2015	21	Undersøkelse 3 A - S1 BKS (14-15)	**Und3 – E**	vg2 - S1	Etter
4	apr 2015	21	Undersøkelse 4 kort test LVS	**Und4 – F**	vg1 - 1T	Før
5	sept 2015	28	Undersøkelse 5 1G BKS(15-16) *	**Und5 – F**	vg1 - 1T	Før
6	jan 2016	22	Undersøkelse 6 1G BKS(15-16) *	**Und6 – E**	vg1 - 1T	Etter
		136				

Tabell. Utvalg til kvantitativ undersøkelse: Gruppe, tidspunkt for undersøkelse, antall elever, beskrivelse av gruppe, navn på gruppe, nivå på gruppe, før eller etter undervisning.

Den første gruppen heter «Pilot» og var riktignok en pilotgruppe. Piloten gikk som smurt, og jeg foretok bare en mindre endring i én av oppgavene etter tilbakemelding fra denne gruppen. Grupper merket (*) er grupper som jeg underviste selv. I kapittel 4 Resultat, vil jeg se på det som jeg kaller for «direkte sammenligning», og det som jeg kaller for «alle grupper». Direkte sammenligning er et utvalg av de syv gruppene ovenfor. Jeg ser da bare på de to første og de to siste gruppene; Sammenligning 1 (Sam1) og Sammenligning 2 (Sam2).

	Før undervisning	Etter undervisning
Sam1	**Pilot – F**	**Und1 – E**
Sam2	**Und5 – F**	**Und6 – E**

De to første gruppene (Sam1) har stor grad av sammenheng, selv om det er to ulike klasser. Gruppene fulgte det samme opplegget hele skoleåret og hadde blant annet de samme prøvene. Gruppene lå parallelt på timeplanen og hadde to ulike lærere. Begge lærerne karakteriserte de to gruppene som «ganske» like. Likheten er ikke dokumentert på annen måte enn at det ved slutten av skoleåret viste seg at gjennomsnittet til standpunktkarakterene for gruppene kun hadde et mindre avvik, med fordel Pilot – F.

De to siste gruppene (Sam2) er samme klasse før og etter undervisning. For å ivareta ulikheten mellom «før» og «etter» ble tidsrommet mellom de to undersøkelsene utvidet. Elevene beklaget seg også over at de dessverre (for dem) ikke husket noe fra forrige undersøkelse. Jeg tror ikke at den første undersøkelsen påvirket den andre undersøkelsen i nevneverdig grad, men sikker kan jeg ikke være. Læreren for denne klassen er meg selv.

Når jeg ikke har direkte sammenligning er alle syv gruppene med, unntaket er noen ganger Und4. De hadde en forkortet versjon av testen og er derfor ikke med på alle oppgavene.

Til den kvalitative delen av oppgaven, som er intervju, trengte jeg hensiktsmessige utvalg. Siden antall individer er få, 10, må utvalget ha mindre grad av tilfeldighet enn utvalget til kvantitative undersøkelser. Elevene jeg har er stort sett engasjerte og interesserte elever. Jeg tilbød derfor timer med ekstraundervisning i matematikk i bytte mot intervju. Det var mange elever som meldte sin interesse, og jeg gjorde et utvalg på tre grupper, hver med tre personer. Utvalgene er valgt fordi de hensiktsmessige, men samtidig er de valgt av behagelighetshensyn. Den siste personen i utvalget ble plukket ut på et senere tidspunkt. Jeg har tatt vare på lydfilene digitalt, men de er ikke transkribert.

Elevene til de tre gruppeintervjuene (1-3) er hentet fra den kvantitative undersøkelsen. Eleven i «en-til-en»-intervjuet (4) tilhører ingen av gruppene. Denne eleven er en gutt som har hatt sannsynlighetsregning på vg2 – S1. Alle intervjuene er gjort *etter* at elevene har gjennomført CPR-testen og hatt undervisning i emnet.

Intervju	**Antall**	**Beskrivelse – internt bruk**	**Navn**	**Nivå**
1	3	Et utvalg fra Pilot – 1F BKS (14-15)	**Intervju 1**	1T
2	3	Et utvalg fra Undersøkelse 1 1G BKS(14-15)	**Intervju 2**	1T
3	3	Et utvalg fra Undersøkelse 1 1G BKS(14-15)	**Intervju 3**	1T
4	1	Gutt – tok CPR-testen like før intervjuet.	**Intervju 4**	S1

Tabell. Utvalg til kvalitativ undersøkelse: Intervju gruppe, antall elever, beskrivelse av gruppe, navn på gruppe, nivå på gruppe. Alle er etter undervisning.

Elevene ble informert om det generelle formålet med oppgaven, men ikke i detaljert grad. De ble informert om at de til enhver tid kunne trekke seg fra intervjuet. Integriteten til elevene er ivaretatt ved at alle elevene er gitt nummer på spørreundersøkelsen og ingen navn blir nevnt i intervjuene.

3.3.3 Intervju

Intervjuene i denne oppgaven er alle med meg til stede. Jeg vil derfor konsentrere meg om de to hovedinndelingene for slike intervju som er

(1) «*en-til-en*»
(2) *fokusgruppe*.

I et «en-til-en»-intervju er det forskeren som intervjuer en enkelt deltaker. Dette er tidkrevende og dermed kostnadskrevende. Det er vanlig å bruke flere «en-til-en»-intervjuer i samme oppgave, men jeg hadde bare ett. Denne metoden passer ypperlig med deltakere som er verbale, kommunikative og som lett kan dele sine ideer.

Fokusgrupper kan brukes til å finne en samlet forståelse av et fenomen. Det typiske antallet er fire til seks personer. Forskeren spør et lite antall generelle spørsmål og samler responsen fra alle. Denne metoden har sin styrke når deltakerne samarbeider og i stor grad består av likeverdige partnere som samarbeider godt. Når tiden til å samle data er begrenset, er bruk av fokusgrupper nyttig. I en fokusgruppe må alle deltakere oppfordres til å snakke slik at alle stemmer kommer frem.

Intervjuene er oppgave-basert fordi de tar utgangspunkt i oppgavene fra CPR og faller inn under definisjonen til Goldin (side 519 i Goldin, 2000). Som et forskningsbasert verktøy reiser oppgave-basert intervju flere spørsmål (side 303 i Goldin, 1993), som jeg senere vil prøve å besvare:

1) Hvordan får vi studert det vi ønsker å studere? Hvordan kan vi gjenta, eller reprodusere, et lignende forsøk senere? Hvordan sammenligner vi de svarene som er gitt? Er det mulig å generalisere?
2) Hvilken rolle spiller strukturen til intervjuet? Hvordan påvirker strukturen intervjuet?
3) Hvilken innvirkning har den sosiale og psykologiske settingen for resultatet?
4) Er det noen generelle prinsipper i design og konstruksjon for denne typen? Vil disse prinsippene igjen kunne optimere informasjonen som kommer ut av intervjuene?

I oppgave-basert intervju kan vi få til matematiske samtaler. Dette er undersøkende samtaler som har til hensikt å fremme læring. I denne prosessen kan misoppfatninger avdekkes.

Jeg ville sette i gang intervjuene noenlunde tidlig, slik at jeg fikk tid til å analysere svarene opp mot analysen fra den kvantitative delen. For å ta det siste intervjuet først. Dette var et «en-til-en»-intervju. Min deltaker var noe mindre kommunikativ enn hva jeg på forhånd hadde tenkt, og nokså sikker på sine egne prestasjoner. Dette førte til at intervjuet ble kort og lite frodig. Intervjuet var semi-strukturert etter følgende modell:

1 Eleven gjennomfører testen før intervjuet

2 Eleven presenterer sitt svar – og blir møtt med et annet alternativ

3 Eleven vurderer de to alternativene opp mot hverandre og tar et valg med begrunnelse

Det alternative svaret som eleven presenteres for under punkt 2, er enten det riktige svaret dersom eleven har svart feil, eller det feilaktige svaret med størst frekvens fra CPR-testen. De andre intervjuene var tre intervju med fokusgrupper. Disse intervjuene var mer åpne i sin form, og jeg hadde i utgangspunktet tenkt å holde meg mer i bakgrunnen for å la elevene lede fremdriften. Dette viste seg nokså vanskelig fordi de lett lot seg lede vekk fra hovedsporet, og jeg måtte ofte trø til for å lede dem på rett vei. Dersom jeg ikke hadde ledet dem, ville tiden som gikk med til hvert intervju vært formidabel og risikoen for en grundig avsporing ville absolutt vært til stede. Det viste seg nemlig at elevene ble leie og trøtte da vi bikket halvtimen. Fokusgrupper har etter teorien gjerne fire til seks personer, men jeg synes at det var mer enn nok med tre i hver gruppe. Blant annet var det utfordrende å skille elevene fra hverandre når lydfilene skulle transkriberes.

Her kommer svarene:

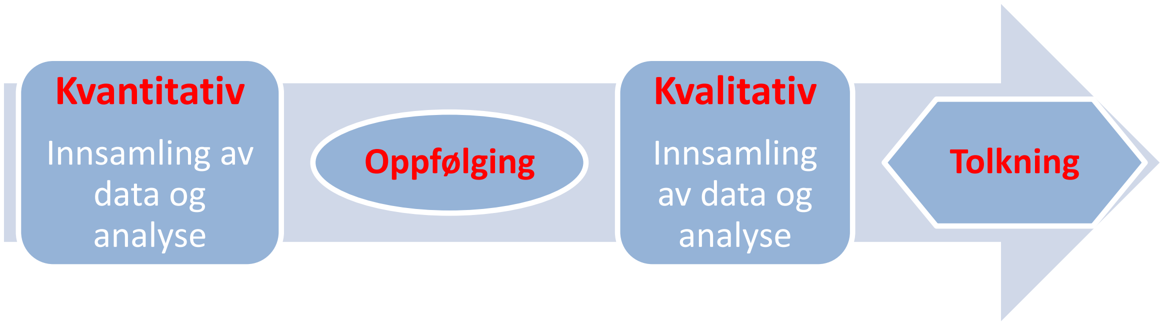
1)	Hvordan får vi studert det vi ønsker å studere? Ved å være deltagende i intervjuet kan vi lede elevene i den retningen vi ønsker.
	Hvordan kan vi gjenta, eller reprodusere, et lignende forsøk senere? Se sammenhengen med spørsmål 2
	Hvordan sammenligner vi de svarene som er gitt? Er det mulig å generalisere? I et åpent intervju kan dette være vanskelig, men ikke umulig. Svarene må tolkes og analyseres i et generelt perspektiv for å kunne sammenlignes.
2)	Hvilken rolle spiller strukturen til intervjuet? Hvordan påvirker strukturen intervjuet? For å kunne reprodusere intervjuet, må intervjuet ha en viss struktur. Som eksempel er intervju 4 lettere å reprodusere enn de andre, fordi dette intervjuet var semi-strukturert.
3)	Hvilken innvirkning har den sosiale og psykologiske settingen for resultatet? Vanskelig å svare på. Alle elevene er rimelig verbale og trygge på hverandre og meg.
4)	Er det noen generelle prinsipper i design og konstruksjon for denne typen? Vil disse prinsippene igjen kunne optimere informasjonen som kommer ut av intervjuene? Optimering av informasjon var vanskelig på grunn av tiden som gikk med til hvert intervju. Som tidligere nevnt ble elevene slitne – og vel en ½ time var nok. Noen oppgaver var så spesielle i sin design at det tok tid å lede elevene inn på riktig spor.

Tabell. Spørsmål og svar til oppgave-basert intervju (side 303 i Goldin, 1993).

Jeg hadde oppgave-basert intervju. Jeg burde i større grad konsentrert meg om et utvalg av oppgavene og da bestemt hvilke oppgaver jeg ville undersøke spesielt. Nå ble det hele noe tilfeldig. For en eventuell reproduksjon av den kvalitative delen av undersøkelsen, bør det rettes fokus mot spesielle oppgaver, og intervjuene bør ha en viss grad av struktur. Det er 13 oppgaver med totalt 18 svar. I intervjusituasjonene oppdaget jeg at elevene ble nokså utslitt etter vel 30 minutter samtale, lenge før vi var ferdige med alle oppgavene. Det var derfor ikke mulig å få til en diskusjon/samtale om hver oppgave i intervjuene. Når jeg vurderer intervjuene i ettertid, ser jeg at jeg burde hatt en pilot på intervjuet i stedet for en pilot på CPR-testen. Intervjuene i fokus gruppene bar preg av matematisk samtale. Min oppfatning er at elevene i samtalene i stor grad avdekket sine misoppfatninger.

3.3.4 Mitt design

Metoden er mixed method, men den legger større vekt på den kvantitative delen av blandet metode enn den kvalitative. Metoden er sekvensiell ved at den kvalitative delen kommer etter den kvantitative. Jeg bruker den kvalitative delen til å undersøke og forklare svarene og analysen fra den kvantitative delen. Designet som prioriterer kvantitative data, er en blandet metode som heter «forklarende sekvensiell design».



Figur: Type blandet metode; Forklarende sekvensiell design (side 541 i Creswell, 2012)

Kapittel 4 Resultat legger vekt på den kvantitative analysen og kapittel 5 Drøfting implementerer den kvalitative analysen i lys av resultatene fra kapittel 4.

3.3.5 Oppgavene

På grunn av at oppgavene i min test ikke er identiske med oppgavene i CPR, er det noen spørsmål som må besvares. CPR-testen er inndelt i små enheter, slik at det er mulig å korte ned testen. Jeg kortet ned testen fordi jeg ville at mengden skulle være overkommelig og at vanskelighetsgraden ikke skulle få en negativ innvirkning. Jeg fryktet at en del av de bayesiske problemene ville virke demotiverende på elever som akkurat har begynt i videregående skole. Jeg ville unngå en overfladisk tilnærmelse og dermed større mulighet for system 1-feil. Oppgavene som er fjernet er i sin helhet gjengitt i kapittel 7 Appendiks.

CPR	Enhet	Begrunnelse
2	3: Løse et bayesisk problem	Stor grad av gjetting på grunn av tidkrevende oppg.
3	7: Identifisere uavhengige og gjensidig ekskluderende hendinger	Valg mellom regneregler og symbolbruk i alt c og d, som elever på vg1, før undervisning, ikke har forutsetninger for å kjenne til.
5	11: Beregne betinget sannsynlighet i et forsøk med union og sammensatte hendinger 15: Beregne betinget sannsynlighet i en synkron situasjon	Elever uten undervisning i betinget sannsynlighet og total sannsynlighet har liten sjanse for å løse oppgaven. Dessuten er brystkreft et litt ømtålig tema. Enhet 15 måles på andre oppgaver.
11	1: Definere betinget sannsynlighet og gi et passende eksempel	Mange elever kjenner ikke til begrepet før undervisning. Totalt ukjent og vanskelig som første åpne oppgave.
14	15: Beregne betinget sannsynlighet i en synkron situasjon	Total sannsynlighet er ukjent. Samtidig måles enhet 15 i andre oppgaver.

Tabell. Oppgaver som er fjernet: Oppgavenummer i CPR. Enhet. Begrunnelse for fjerning.

For noen oppgaver skiller poengene fra CPR seg fra mine undersøkelser. CPR velger å gi 2 poeng for riktig svar på de åpne oppgavene og 1 poeng for delvis riktige svar. Samtidig velger CPR å slå sammen oppgave 8a og 8b til én oppgave 8. Jeg har valgt å gi alle oppgaver 1 poeng, rett og slett for enkeltheten til oppgaven. Å endre dette i mine excel regneark er ingen større operasjon, siden jeg har notert antall delvis riktige svar også. Jeg er enig i at det kan være viktig å identifisere delvis riktige svar, men jeg synes at det er for mye at de åpne oppgavene skal telle dobbelt av multiple choice-oppgavene, eller at delvis riktig svar (som ofte var langt fra riktig) skal telle like mye som et riktig svar på multiple choice.

3.4 Kontroll

Jeg har gjennomført mange ulike undersøkelser i denne masteroppgaven. Kvaliteten på en oppgave er avhengig av gjennomføringen av undersøkelsene og valg av type undersøkelse. Dette beskrives best ved reliabilitet, vurdering av gjennomføring av metode, og validitet, vurdering av valg av metode.

3.4.1 Reliabilitet

Reliabiliteten til en oppgave er en del av kvalitetskontrollen. Er oppgaven og resultatet fra denne å stole på? Hvor konsistent eller nøyaktig er det resultatet vi får? Er oppgaven troverdig?

- Graden av pålitelighet og nøyaktighet er tegn på reliabilitet.
- Graden av reliabilitet er tegn på kvalitet.

God reliabilitet betyr at data er lite påvirket av tilfeldige målingsfeil. Koding og analyse av alle data er utført av meg. Tolkninger av transkripsjon i intervjudelen er gjort meg. Jeg vil derfor si at reliabiliteten er god, eller i alle fall nokså ensrettet, siden det ikke er andre personavhengige faktorer.

Går elevene lei? Tar det for lang tid? Dette kan være faktorer som påvirker reliabiliteten til en oppgave, og var også en medvirkende årsak til at jeg kortet ned på antall oppgaver. Er figurene gode nok. Er teksten god nok? CPR-testen er gjennomgått grundig av flere forskere slik det er beskrevet i kapittel 3.2. Jeg gjorde kun en mindre endring i teksten, etter råd fra veileder. Svarene er analysert med prosentfordeling på hvert svaralternativ. Det er regnet ut et gjennomsnitt for hver gruppe. Det er ikke beregnet korrelasjon, signifikans eller tatt andre tester for å studere reliabiliteten til resultatet. Hver gang oppgaven er gjennomført er det gitt like lang tid til elevene. Ingen elever har fått hjelp underveis. Elevene kunne bruke lommeregner underveis, men det var ikke nødvendig med utregninger. Et riktig oppsatt regnestykke talte like mye. Oppgavene og oppgaveteksten er den samme for alle. Jeg vil derfor påstå at denne delen av masteroppgaven har god reliabilitet, fordi den kan gjentas - og har blitt gjentatt på en annen skole, av en annen person til en annen tid – uten problemer av noe slag. Reliabiliteten til hver enkelt oppgave kan diskuteres fordi de ikke er rotert i rekkefølge. Det kan føre til lavere score på de siste oppgavene fordi deltakerne går lei under testen. Dette er en god grunn til å korte ned på antall oppgaver fra den opprinnelige CPR-testen. Jeg vil ut fra dette si at reliabiliteten i den oppgavebaserte delen er god.

Jeg har utført intervju tidligere, men var absolutt i ukjent terreng denne gangen fordi jeg ikke visste hvordan en fokusgruppe ville gripe an situasjonen. Intervjuene var derfor i liten grad strukturert. Som tidligere nevnt, burde jeg kanskje ha satt inn piloten her i stedet for på den kvantitative delen av masteroppgaven. Det er altså grunn til å diskutere reliabiliteten i intervjuene. Utformingen og vanskelighetsgraden til oppgavene gjorde det naturlig til at elevene diskuterte seg frem til en løsning. Kvaliteten på intervjuet er avhengig av den som utfører intervjuet. Jeg ledet dem ofte på veien med positive/negative «Hmm». Et «Hmm» kan synges bekreftende (positiv), som om du er på riktig vei, eller den kan synges avkreftende (negativt): «Hmm (skeptisk, er du nå sikker på dette da?)», uten at det siste blir sagt, ligger det implisitt. Jeg har også mange innspill med «Ja» og noen «Nei». Jeg har i tillegg hjulpet elevene til å omformulere det de sier, slik at det kan lede de andre deltakerne inn på den samme og den riktige veien, mot målet – og det riktige svaret. Kandidatene som var med på intervju gjorde sitt beste for å delta i diskusjonen, selv om ikke alle var like kommunikative. Jeg er av den oppfatning at min medvirkning var nødvendig for fremdriften og for å finne svar, men samtidig vil de ustrukturerte intervjuene ha lav reliabilitet, blant annet fordi de vil være vanskelig å gjenskape med en annen intervjuer på et annet tidspunkt.

Intervjuet med en enkelt elev er lettere å gjenskape fra intervjusiden, men mindre lett fra elevsiden. Under slike intervju er man til stor grad prisgitt eleven som er til intervju. Se min beskrivelse av kandidat under kapittel 3.3. Jeg vil derfor si at semi-strukturert enkeltintervju ikke er noen garanti for reliabilitet, og at det er grunn til å stille spørsmål vedrørende reliabiliteten i den kvalitative delen.

Under transkripsjon av intervjuene har jeg i mindre grad tatt hensyn til stemmeleie og uttrykk som latter, frustrasjon og lettelse. Jeg har latt teksten stort sett stå ren og i liten grad sett på kroppsspråk. Noen ganger var det vanskelig å høre eksakt hva som ble sagt og noen ganger var det vanskelig å skille kandidatene fra hverandre. Jeg vil likevel si at reliabiliteten på transkripsjonen er god.

3.4.2 Validitet

Validiteten er sammen med reliabilitet en del av kvalitetskontrollen. Hvor godt måler min undersøkelse det den er ment å måle?

- Graden av gyldighet og relevans er tegn på validitet
- Graden av validitet er tegn på kvalitet

Jeg må søke etter gode beviser for at min metode er relevant og gyldig. Er mixed methods og gjennomføringen av oppgaver og intervju et godt valg for det jeg ønsker å si noe om? At en test eller et instrument har ytre validitet avhenger av i hvor stor grad resultatet i testen kan overføres fra utvalgene til populasjonen i sin helhet. Presterer elevene i mine utvalg bedre etter undervisning?

Spørreskjema med oppgaver er todelt med multiple choice og åpne oppgaver. De første oppgavene er konsistente, men gjennomgår til dels en drøfting i kapittel 5. De åpne oppgavene er i mindre grad konsistente, men artikkelforfatterne har gitt en grundig beskrivelse av alle svar, slik at tolkningen av riktige, delvis riktige og feilaktige svar er lik for alle gruppene i min oppgave og samtidig lik tolkningen til Diaz med flere (Diaz og Batanero, 2009). Der det er eventuelle uoverensstemmelser mellom Diaz med flere og meg, er dette beskrevet i analysen av hver oppgave. Jeg mener at validiteten i denne delen av oppgaven er god, fordi jeg kan, i mer eller mindre grad, hake av alle sjekkpunktene for hva som skal og bør være med (side 162 i Creswell, 2012)

- ✓ Finn et instrument (spørreskjema) du vil bruke
- ✓ Se etter om andre har brukt tilsvarende instrument tidligere
- ✓ Se nøye på formålet for instrumentet i tidligere undersøkelser
- ✓ Se nøye på tolkningen av svarene instrumentet gav
- ✓ Vurder om forfatterne gir gode bevis for sin nytte av denne tolkningen

Det er noen begrensninger til validiteten som for eksempel at utvalget neppe oppfyller kravene eller definisjonene til et tilfeldig utvalg. For eksempel har så å si alle elevene samme lærer: meg. Resultatet er derfor ikke allmenngyldig. For øvrig bekreftes validiteten til CPR-testen for Diaz med flere i kapittel 4.14, mens jeg drøfter den tilsvarende validiteten i mine undersøkelser i kapittel 4.16.

Selv om jeg var dårlig forberedt på intervjusituasjonen og kanskje hadde en mulighet til å få mer ut av situasjonen, er kanskje funnene jeg gjorde generaliserbare? Jeg vil vise til funn som harmonerer med funnene i CPR-testen for Diaz med flere og nye funn som ikke er omtalt tidligere. Intervjuene avslører samtidig at mange av funnene i den kvantitative undersøkelsen er generaliserbare. Jeg har vært kritisk til intervjuene og min rolle, men jeg vil samtidig påpeke at rollen og formen på intervjuene, med ledende spørsmål og tilnærmet rettledning, også har hjulpet til å avdekke misoppfatninger omtalt i kapittel 2. Jeg er av den oppfatning at min medvirkning var nødvendig, men samtidig vil de ustrukturerte intervjuene ha lav reliabilitet blant annet fordi de vil være vanskelig å gjenskape med en annen intervjuer på et annet tidspunkt. Resultatet fra intervjuene og sitater i kapittel 5 vil vise god validitet.

3.5 En utfordring

Denne undersøkelsen har flere nummereringer på kryss og tvers. For de to neste kapitlene vil jeg presentere noen av utfordringene som dukker opp. Oppgavene til den kvantitative delen av undersøkelsen er nummerert fra 1 -13. For å unngå videre forviklinger, vil jeg holde meg til min nummerering på oppgavene jeg tok med, og utelate oppgavenummer fra CPR.
I CPR-testen til Diaz med flere er det klassifisert ulike enheter innen betinget sannsynlighet, nummerert fra 1 til 19. Jeg vil forholde meg til de samme numrene, selv om noen av enhetene faller ut fordi jeg har fjernet noen oppgaver. Tabellene nedenfor viser enhetene.

Oppgave	**Enheter fra CPR**	
1a	4	Identifisere betingede, **enkle** og sammensatte sannsynligheter
1b	4	Identifisere betingede, enkle og **sammensatte** sannsynligheter
1c	4	Identifisere **betingede,** enkle og sammensatte sannsynligheter
	15	Beregne betinget sannsynlighet i en synkron situasjon
1d	4	Identifisere **betingede,** enkle og sammensatte sannsynligheter
	15	Beregne betinget sannsynlighet i en synkron situasjon
2	10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging
	14	Beregne betinget sannsynlighet i en asynkron situasjon
3	6	Feilslutningen om snitt i sammensatte hendinger (conjuction fallacy)
4	5	Identifisere en betinget sannsynlighet og den inverse sannsynligheten
	13	Identifisere betinget årsak- og diagnostisk situasjon
5	12	Beregne betinget sannsynlighet i et forsøk med omvendt tidsakse
	14	Beregne betinget sannsynlighet i en asynkron situasjon
6a	10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging
	14	Beregne betinget sannsynlighet i en asynkron situasjon
6b	10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging
	12	Beregne betinget sannsynlighet i et forsøk med omvendt tidsakse
	14	Beregne betinget sannsynlighet i en asynkron situasjon
7	14	Beregne betinget sannsynlighet i en asynkron situasjon
	17	Beregne sannsynlighet med produktsetningen på to avhengige hendinger
8a	2	Gjenkjenne at en betingelse gir en begrensning av utfallsromet
8b		kombinatoriske evner
9	8	Beregne betinget sannsynlighet i ett enkelt forsøk
10	9	Beregne betinget sannsynlighet i et forsøk med tilbakelegging
11	16	Beregne sannsynlighet med produktsetningen på to uavhengige hendinger
12	15	Beregne betinget sannsynlighet i en synkron situasjon
	17	Beregne sannsynlighet med produktsetningen på to avhengige hendinger
13	19	Løse bayesisk problem

Tabell fra kap 3.2. Oppgavenummer fra min undersøkelse. Enhet nummer og beskrivelse.

Oppgavene kan ha forbindelse til flere enheter enn det som er vist i tabellen ovenfor.
Tabellen nedenfor viser sammenhengen mellom misoppfatningene, nummerert fra 1 til 8, slik de er beskrevet i kapittel 2, og oppgavene som i hovedsak hører til misoppfatningene.

Nr	**Misoppfatning fra kapittel 2**	**Oppgave**
1	Kontekst og forveksling	1
2	Snittet mellom to hendinger, conjunction fallacy	3
3A	Motsatte betingede sannsynligheter - forveksling og likhet	4
3B	Motsatte betingede sannsynligheter - inversjon av tidsaksen	5 og 6b
3C	Motsatte betingede sannsynligheter - bayesiske problemer	13
4	Synkrone og asynkrone hendinger	2, 6a, 7, 11, 12
5	Heuristikker - Siste observasjoner og representativitet	10
6A	Den mest vanlige - bestemt sannsynlighet	9 og 10
6B	Den mest vanlige - lik sannsynlighet	9 og 10
6C	Den mest vanlige - bestemt regel	9
7	Kombinatoriske problem	8
8A	Heuristikker - tilgjengelige utfall	
8B	Heuristikker - utfallstilnærming	
8C	Heuristikker - Base Rate	
Syst1	System 1-feil. Beskrevet i kapittel 1	9

Tabellen. Misoppfatning og tilhørende nummer i denne oppgaven.

Det kan være flere oppgaver som kan knyttes til samme misoppfatning enn det som er vist i tabellen ovenfor. Dette blir diskutert under videre drøfting i kapittel 5. Vi ser også at det ikke er noen oppgaver som direkte er linket til misoppfatning nummer 8. Disse heuristikkene blir også drøftet i kapittel 5.

4 Resultat

I dette kapitlet vil jeg sammenligne prestasjonene til de ulike gruppene før og etter undervisning. Vil prestasjonen bli bedre med undervisning? Hovedproblemstillingen og underspørsmål:

Hvilken innvirkning har undervisning i skolen på misoppfatninger innen betinget sannsynlighet?

Vil undervisning redusere antall misoppfatninger?

Hvilke misoppfatninger består, øker eller minker etter undervisning?

Vi trenger en liten presisering av sammenhengen mellom problemstillingen og innholdet i dette kapitlet. Hovedfokus er resultatet i hver oppgave. Jeg studerer i hvilken grad mine resultater harmonerer med resultatene til Diaz med flere (Diaz og Fuente, 2007/Diaz og Batanero, 2009). Spørsmålet om vi kan konkludere med at misoppfatningene reduseres dersom resultatet viser til en forbedring blir besvart via drøfting i kapittel 5. Kapittel 4 er altså mer resultatorientert enn hva hovedproblemstillingen kanskje skulle tilsi. For enkelthets skyld gjentar jeg gruppenavnene fra mine undersøkelser slik det ble presentert i kapittel 3. Det er navnene, eller deler av disse, i kolonnen med fet skrift som blir brukt i videre presentasjonene av resultatene.

Gruppe	Tidspunkt	Antall	Beskrivelse – internt bruk	Navn	Nivå	Før/etter
Pilot	jan 2015	16	Pilot – 1F BKS (14-15)	**Pilot – F**	vg1 - 1T	Før
1	mars 2015	14	Undersøkelse 1 1G BKS(14-15)	**Und1 – E**	vg1 - 1T	Etter
2	mars 2015	14	Undersøkelse 2 M - S1 BKS (14-15)	**Und2 – E**	vg2 - S1	Etter
3	mars 2015	21	Undersøkelse 3 A - S1 BKS (14-15)	**Und3 – E**	vg2 - S1	Etter
4	apr 2015	21	Undersøkelse 4 kort test LVS	**Und4 – F**	vg1 - 1T	Før
5	sept 2015	28	Undersøkelse 5 1G BKS(15-16)	**Und5 – F**	vg1 - 1T	Før
6	jan 2016	22	Undersøkelse 6 1G BKS(15-16)	**Und6 – E**	vg1 - 1T	Etter
		136				

Tabell. Utvalg til kvantitativ undersøkelse: Gruppe, tidspunkt for undersøkelse, antall elever, beskrivelse av gruppe, navn på gruppe, nivå på gruppe, før eller etter undervisning.

Når jeg skal sammenligne resultatene vil jeg velge fritt hvilke grupper jeg tar med og utelater. For eksempel vil jeg ofte utelate S1-gruppene fordi disse to gruppene noen ganger vanner ut spesielle inntrykk eller virker forstyrrende på andre måter. Jeg kan velge å finne gjennomsnittlig score for alle 10 (9) gruppene eller å separere dette i «før» og «etter» - grupper. Jeg vil også bruke det jeg kaller direkte sammenligning, der enda flere grupper er utelatt og kun spesielle grupper er med.
Ved en slik fremgangsmåte og valg fra øverste hylle kan det virke som om jeg bruker og utelater tallene slik det passer meg, slik at jeg får de svarene jeg vil ha – og den kritikken er nok riktig. På den andre siden forteller jeg om hva jeg gjør hver gang jeg gjør det. Jeg bruker denne fremgangsmåten fordi jeg føler at å låse seg til en bestemt oppskrift eller metode i større grad vil være et hinder fremfor en styrke. Jeg åpner for muligheten til å se på et utvalg fremfor hele populasjonen. Min populasjon kan være et mindre representativt utvalg enn mine utvalg. Spørsmålet er hvem som representerer den totale populasjonen best? Svaret vites ei, men det kan absolutt drøftes. Vi har:

Sam1: direkte sammenligning mellom Pilot (før) og Und1 (etter)

Sam2: direkte sammenligning mellom Und5 (før) og Und6 (etter)

Jeg skal sammenligne mine resultat med resultat fra Diaz med flere som opererer med følgende navn

D & B – F	Batanero og Diaz	Før
D & B – E	Batanero og Diaz	Etter
D & F – E	Diaz og Fuente	Etter

Jeg går rett på resultatene for alle grupper i de 13 oppgavene. I min analyse av resultatet vil jeg ofte bruke forkortingen «Diaz med flere» uten referanse til litteraturliste. Til slutt viser jeg samlet resultat.

4.1 Resultat oppgave 1

Denne oppgaven består i å lese sannsynligheter ut fra en krysstabell. Spørsmålene er beregning av enkel, sammensatt og betinget sannsynlighet. Oppgavetypen er nokså vanlig på vg1T-nivå. Oppgaven er kategorisert som en multiple choice-oppgave i CPR, trolig fordi vi finner alle nødvendige data i tabellen. Definisjon av hendingene H og E og svarene er føyet til av meg.

Oppgave 1

I en medisinsk undersøkelse ble en gruppe personer intervjuet med følgende resultat:

	55 år eller yngre	Eldre enn 55 år (E)	Total
Tidligere hjertestans (H)	**29**	**75**	**104**
Ikke tidligere hjertestans	**401**	**275**	**676**
Total	**430**	**350**	**780**

Anta at vi velger en tilfeldig person fra denne gruppen:

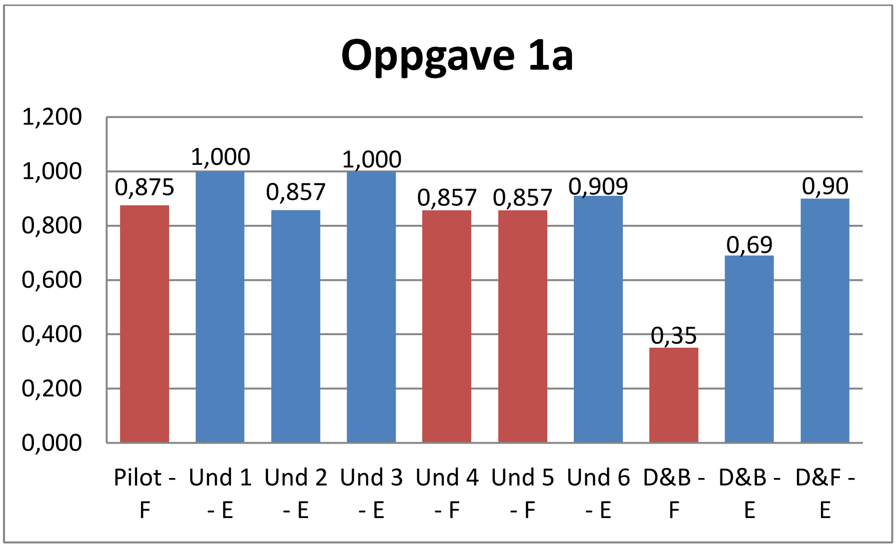
a) Hvor stor er sannsynligheten for at personen har hatt hjertestans? $P(H)=\frac{104}{780}$

b) Hvor stor er sannsynligheten for at personen har hatt hjertestans og samtidig er eldre enn 55 år? $P(H\cap E)=\frac{75}{780}$

c) Dersom en person er eldre enn 55 år, hvor stor er sannsynligheten for at personen har hatt hjertestans? $P(\mathrm{H}|\mathrm{E})=\frac{75}{350}$

d) Dersom en person har hatt hjertestans, hvor stor er sannsynligheten for at personen er eldre enn 55 år? $P(E|H)=\frac{75}{104}$

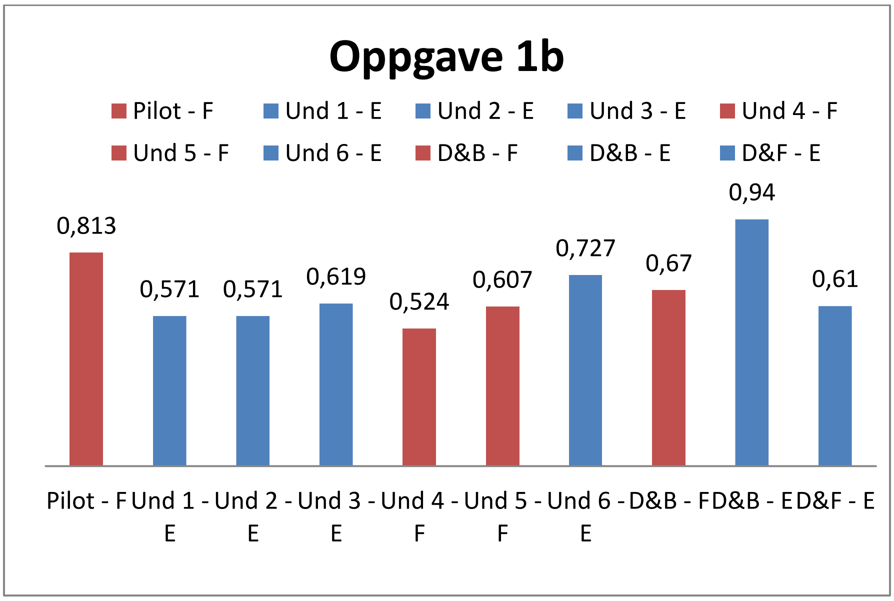
Alle grupper med rød stolpe og bokstav F, henviser til grupper som har tatt testen før undervisning. Grupper med blå stolpe og bokstav E har tatt testen etter undervisning. De tre siste resultatene til høyre i diagramet nedenfor er fra Diaz med flere (Diaz og Fuente, 2007/Diaz og Batanero, 2009). Slik vil det være i alle oppgaver. Oppgave 1a går ut på å identifisere enkel sannsynlighet.



Figur: Oppgave 1a. Prosent riktige svar for hver gruppe. F = Før og E = Etter.

To grupper, som begge har hatt undervisning, får full score på oppgave 1a. Vi ser også at den ene S1-gruppen (Und2), scorer omtrent likt med tre grupper på vg1 – 1T, med relativt flinke elever, som ikke har hatt undervisning. Det er en liten forbedring fra Und5 til Und6 (Sam2), men noe større forbedring fra PilotF til Und1 (Sam1). De tre beste resultatene finner vi i grupper etter undervisning.
I min undersøkelse kan vi si at det er et tegn på at undervisningen hjelper til å oppklare forvekslinger mellom enkle sannsynligheter nevnt under misoppfatning 1 - kontekst. Men på grunn av at prestasjonen allerede er høy for alle mine grupper uten undervisning, blir det litt vanskelig å konkludere. Tegnet på at undervisningen hjelper når man skal lese sannsynligheter ut fra en krysstabell, er mye tydeligere hos Batanero og Diaz. Resultatet viser en økning fra 35 % til 69 % (Diaz og Batanero, 2009).

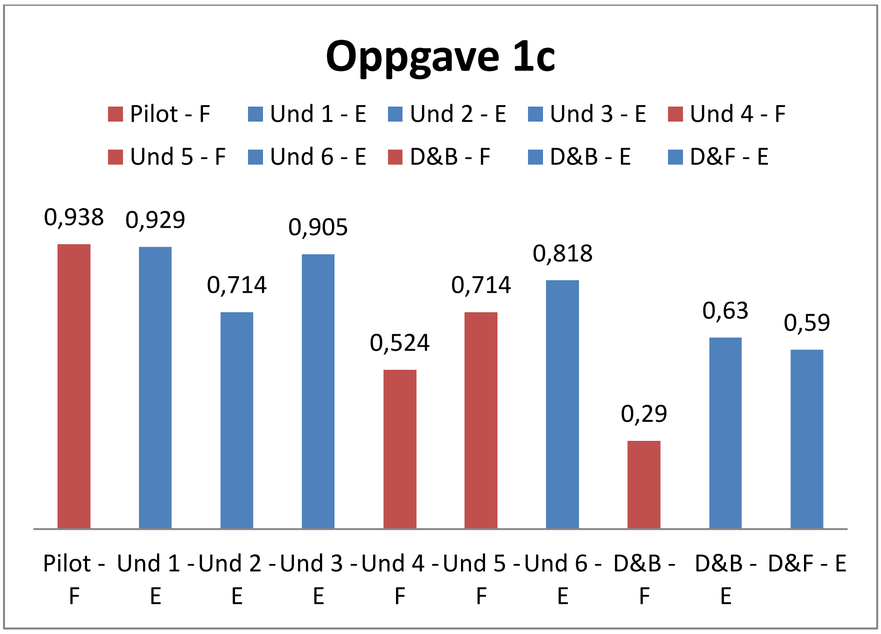
Oppgave 1 b går ut på å identifisere sammensatt sannsynlighet.



Figur: Oppgave 1b. Prosent riktige svar for hver gruppe. F = Før og E = Etter.

Oppgaven gir i min undersøkelse ingen forbedring. Sammenligning mellom Pilot og Und1 (Sam1) viser en klar reduksjon, mens sammenligning mellom Und5 og Und6 (Sam2) viser en moderat forbedring. Vi kan se en klar forbedring mellom de to gruppene til Diaz og Batanero. For de andre gruppene er det lite som skiller, men den dårligste gruppen er «før» og den beste er «etter».
En forsiktig konklusjon kan være at elevene på vg1/vg2 ikke er blitt nevneverdig bedre til å finne sannsynligheten til snittet av to hendinger gitt i en krysstabell, og at de fremdeles kan la seg forvirre av konteksten.

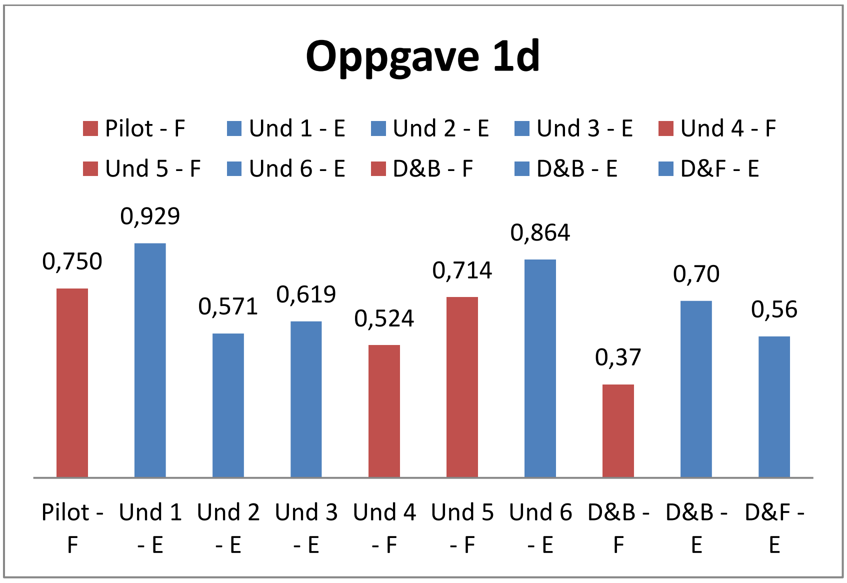
Oppgave 1c og oppgave 1d består i å identifisere betinget sannsynlighet, to omvendt betingede sannsynligheter. Vi ser først på oppgave 1c.



Figur: Oppgave 1c. Prosent riktige svar for hver gruppe. F = Før og E = Etter.

Gruppene som har hatt undervisning scorer høyere, med unntak av den relativt flinke pilotgruppen, som faktisk scorer høyest av alle. Vi ser at elevene i undersøkelse 1, 2 og 3, har høyere eller lik score som elevene i Und4 og Und5. Und6 scorer relativt bedre enn Und5 (Sam2). Det er som kjent en forskjell på vg1T- og S1-elever i prestasjoner, men likevel har Und4 og Und5 (vg1T) lavere eller lik score som Und2 og Und3 (S1). Gruppene til Diaz med flere opplever en markant forbedring. Dette er en indikasjon på at undervisning i betinget sannsynlighet hjelper, og at elevene oppdager begrensningen av utfallsrommet i krysstabellen.

Oppgave 1d består i å identifisere betinget sannsynlighet, men den motsatte av oppgave 1c.



Figur: Oppgave 1d. Prosent riktige svar for hver gruppe. F = Før og E = Etter.

Resultatet mellom Pilot og Und1 viser en markant forbedring. Det gjelder også mellom Und5 og Und6. Vi ser også at S1 (Und2 og Und3) holder noenlunde stand mot vg1T (Und4 og Und5). Vi ser en markant forbedring hos Diaz med flere. Resultatene samsvarer bra med oppgave 1c. Dette er tydelige tegn på at undervisning hjelper å identifisere betinget sannsynlighet i oppgaver med kontekst, som nevnt under misoppfatning 1. Et annet spørsmål er i hvor stor grad elevene etter undervisning forveksler to motsatte betingede sannsynligheter, nevnt under misoppfatning 3A. Dette spørsmålet drøftes i kapittel 5.

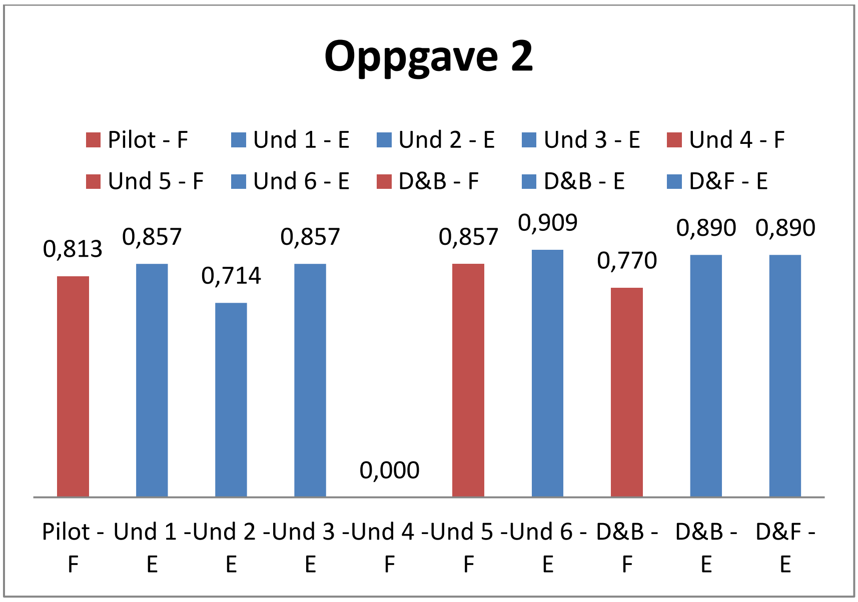
Vi ser at norske elever uten undervisning i emnet, generelt scorer høyt. Forskjellen mellom undervisning eller ikke undervisning er derfor liten. Forskjellen er mer markant for studentene til Diaz med flere. Vi skal også huske på at studentene til Diaz alle hadde gjennomgått undervisning i sannsynlighet på videregående nivå, og at de studentene som er beskrevet som «etter», har hatt spesiell undervisning i sannsynlighet om lag to uker før testen. Tidsrommet mellom generell undervisning og CPR-testen, på opptil to-tre år, kan ses på som grunnlag for den relativt lave scoren «Diaz og Batanero – Før» oppnår. Tilsvarende kan det gode resultatet til «etter-gruppene» tilskrives det relativt korte tidsrommet mellom undervisning og CPR-testen. Misoppfatninger knyttet til kontekst i oppgaver blir redusert med undervisning, men mange forvekslinger mellom enkle, sammensatte og betingede sannsynligheter består.

4.2 Resultat oppgave 2

I alle oppgavene med multiple-choice er det riktige svaret uthevet og fargelagt. Konteksten i denne oppgaven tilsvarer vanlige oppgaver på vg1- og vg2-nivå, men spørsmålsformuleringen er noe ukjent fordi vi i norske oppgaver ofte skal finne og tallfeste en bestemt sannsynlighet. Vi skal beregne betinget sannsynlighet uten tilbakelegg i en asynkron situasjon beskrevet under misoppfatning 4.

Oppgave 2
Det er fire lamper i en boks, to av dem er defekte. Vi plukker ut to tilfeldige lamper fra boksen, den ene etter den andre, uten å legge dem tilbake i boksen. Gitt at den første lampen er defekt; hvilket av alternativene i, ii eller iii er riktig?
Les gjennom alternativene før du svarer. Du må sette alternativene opp mot hverandre.
i) Det er større sjanse for at den andre lampen er defekt.
ii) Det er større sjanse for at den andre lampen virker.
iii) Sannsynlighetene for at den andre lampen er defekt eller at den virker, er like stor.

Det var kun 1 elev av de 136 som deltok som svarte alternativ i. Andelen som svarte alternativ iii, tilsvarer omtrent den resterende delen opp til 100 % for alle gruppene. Gruppen Und4 hadde ikke denne oppgaven i sin forkortede versjon av testen. Jeg har likevel valgt å laUnd4 være representert i alle diagrammer i den tro at det virker mindre forstyrrende for leseren. Diagrammet nedenfor viser også hvordan Und4 blir presentert når de *ikke* har hatt oppgaven.



Figur: Oppgave 2. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Vi observerer generelt høy score for alle gruppene. De er omtrent jevngode, men blant før- og etter-gruppene, er økningen er størst hos Diaz med flere. Jeg vil si at denne oppgaven viser at elevene i stor grad behersker å finne sannsynligheten for asynkrone hendinger intuitivt, og at de oppdager at betingelsen i konteksten begrenser utfallsrommet. Samtidig viser direkte sammenligning mellom Sam1 (0,813 -> 0,857) og Sam2 (0,857 -> 0,909) at undervisning hjelper. Dette kan se ut som tilsynelatende små endringer, men relativt sett er endringene signifikante. Andelen som svarer feil er redusert med 23,5 % og 36,4 % i de to utvalgene.

4.3 Resultat oppgave 3

Etter min erfaring er denne oppgaven utenfor vanlig oppgaveterminologi i den norske skole. Undervisning på vg1 gir en innføring i mengder med snitt og union, men bare i mindre grad. I vg2 brukes venndiagram i forbindelse med total sannsynlighet og dette bør kunne gi i et bedre resultat.

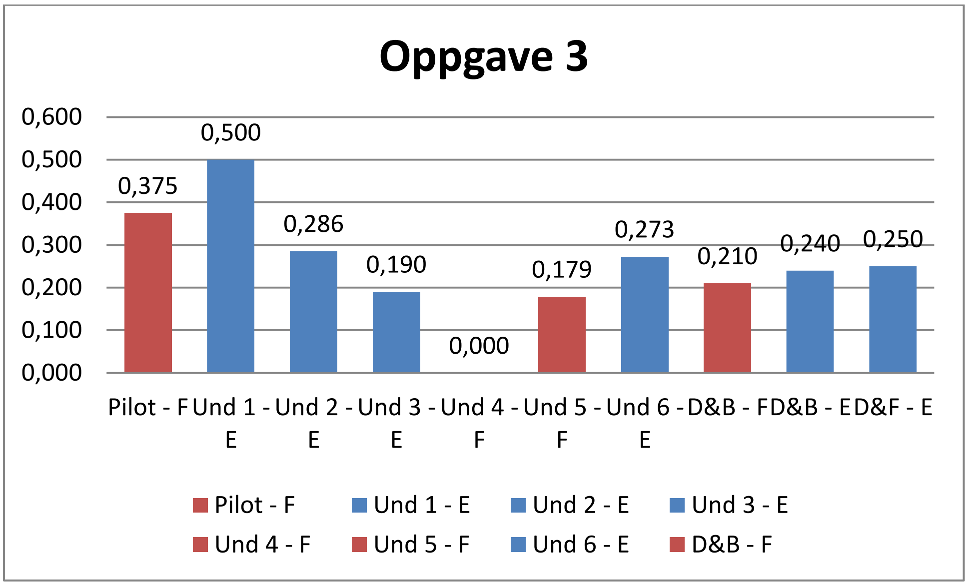
Oppgave 3
En tennisspiller er i finalen i US Open. For å vinne kampen må han vinne tre av fem sett. Hvilke av følgende hendinger er mest sannsynlig eller er de alle like sannsynlige?

a) Spilleren vil vinne det første settet

b) Spilleren vil vinne det første settet, men tape kampen.

c) Begge hendinger a og b er like sannsynlige.

De to feilaktige alternativene tilhører misoppfatning 2, conjunction fallacy, eller misoppfatningen om at den enkelte hending er mindre enn snittet mellom to hendinger. Gruppen Und4 hadde ikke denne oppgaven i sin forkortede versjon av testen.



Figur: Oppgave 3. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Vi ser liten variasjon mellom gruppene i undersøkelsene til Diaz med flere. I mine undersøkelser er variasjonen i prestasjonene større. S1-gruppene (Und2 og Und3) ligger på nivå med Diaz med flere. Det mest interessante er at ingen grupper scorer høyere enn 50 prosent. Direkte sammenligning mellom Pilot og Und1 (Sam1) og Und5 - Und6 (Sam2) viser en økning på 10 – 12 prosent. Undervisning er altså til noe hjelp, men vi må si at den generelt lave scoren viser at conjunction fallacy står sterkt - uansett. Dette harmonerer også med teorien som viste at mange forsøk på endringer og forbedringer av ulike oppgavetekster ikke ga noe bedre resultat (side 56-63 i Tversky og Kahneman, 1974/side 297-300 i Tversky og Kahneman, 1983).

4.4 Resultat oppgave 4

Etter min erfaring er denne oppgaven også noe på siden av det som er vanlig i norsk skole. Elevene skal identifisere to omvendt betingede sannsynligheter. Som nevnt i teorien og misoppfatning 3A, er det mange elever som ikke ser forskjell på alternativ a og b nedenfor. Oppgaven går ut på å identifisere en betinget sannsynlighet og den inverse sannsynligheten.

Oppgave 4

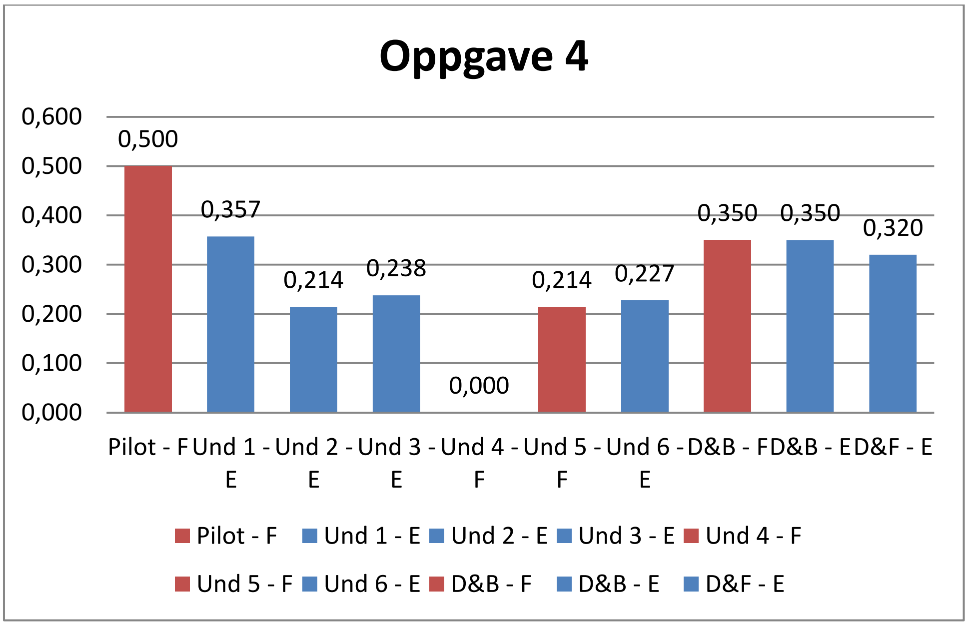
En kreftundersøkelse blir gjennomført på alle innbyggerne i en by. Et positivt resultat på undersøkelsen er en indikator på kreft, mens et negativt resultat indikerer at personen er frisk. Hvilken av følgende resultater er mest sannsynlig eller er de like sannsynlige?

a) Personen har kreft når testen er positiv.

b) Å ha en positiv test når personen har kreft.

c) De to hendingene a og b er like sannsynlige.

Denne oppgaven var ikke en del av testen til Und4.



Figur: Oppgave 4. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Vi ser at de fleste har vanskeligheter med å identifisere de to omvendt betingede sannsynlighetene. Som i oppgave 3, svarer alle grupper lavere enn 50 %, unntatt pilotgruppen som ligger på akkurat 50 %. Hva skal vi tro om svarene?

1) At elevene tror at to omvendt betingede sannsynligheter alltid er like store?
2) At elevene ikke ser forskjell – det er to sider av samme sak?

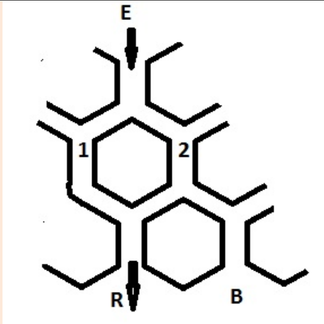
Min tolkning av enhet 5 heller mot det siste alternativet. Med bakgrunn i resultatene kan vi si at undervisning ikke oppklarer denne misoppfatningen. Tvert imot kan vi kanskje si at undervisning forsterker den. Direkte sammenligning mellom Pilot – Und1 (Sam1) viser en nedgang, mens Sam2 og Diaz med flere viser ingen endring. De tre beste blå stolpene mot de tre røde stolpene gir et negativt resultat. Vi kan ikke si at undervisning hjelper.

4.5 Resultat oppgave 5

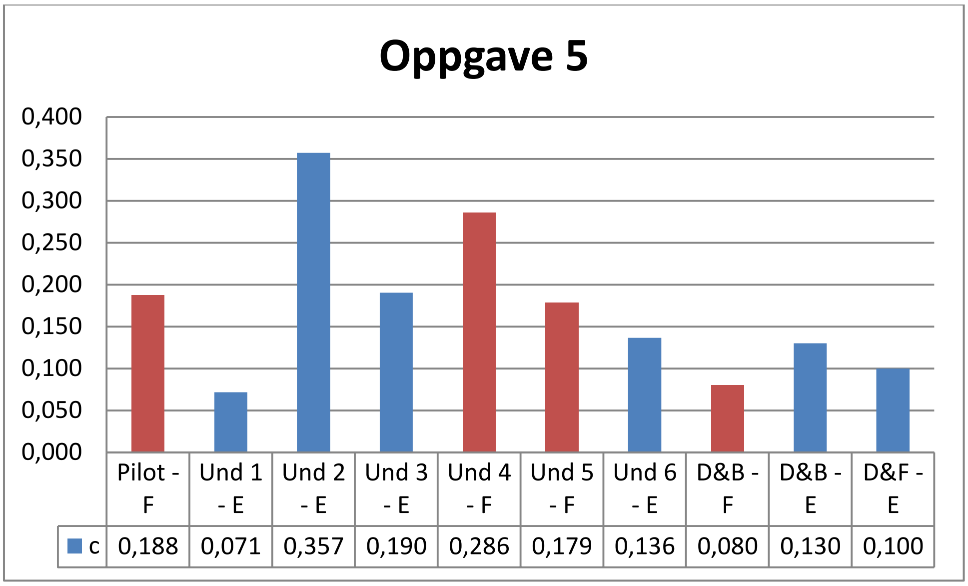
I denne oppgaven skal vi beregne betinget sannsynlighet med omvendt tidsakse. Det er misoppfatning 3B, inversjon av tidsaksen, som er knyttet til denne oppgaven. Det ligger implisitt i begrepet «tidsakse» at hendingene skjer i en rekkefølge og dermed er asynkrone.

Oppgave 5
Vi slipper en ball inn gjennom åpning E på toppen i en maskin som vist på figuren. Anta at ballen forlater systemet gjennom utgang R, hvor stor er sannsynligheten for at den passerer gjennom kanal 1?

a) ½
b) 1/3
c) 2/3
d) umulig å avgjøre



Resultatene i de ulike gruppene er vist nedenfor:



Figur: Oppgave 5. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Først observerer vi at alle gruppene ligger under 40 prosent riktig svarandel. Vi ser også at gruppene med lavest score (Und1 = 0,071) og høyest score (Und2 = 0,357), begge har hatt undervisning. Selv om den visuelle størrelsen til stolpene i diagrammet kan virke dramatiske, er ikke forskjellene så store. Variasjonsbredden er 0,357 – 0,071 = 0,286.

Direkte sammenligning for mine grupper gir en negativ endring av prestasjonen. Pilot presterer bedre enn Und1 og Und5 presterer bedre enn Und6. For Diaz med flere er det en svak positiv tendens i prestasjonen, men resultatet er samtidig lavt. Samlet resultat viser at elever ikke presterer bedre etter undervisning, snarere tvert imot. Undervisning er til hinder for tenkning med omvendt tidsakse, og misoppfatningen består – og kanskje øker? Det er også interessant å merke seg at de fem beste resultatene er blant de litt svakere S1-gruppene og 1T før undervisning. Dette er mer enn et interessant resultat, snarere oppsiktsvekkende. Dette krever drøfting i både kapittel 5 og 6.

4.6 Resultat oppgave 6

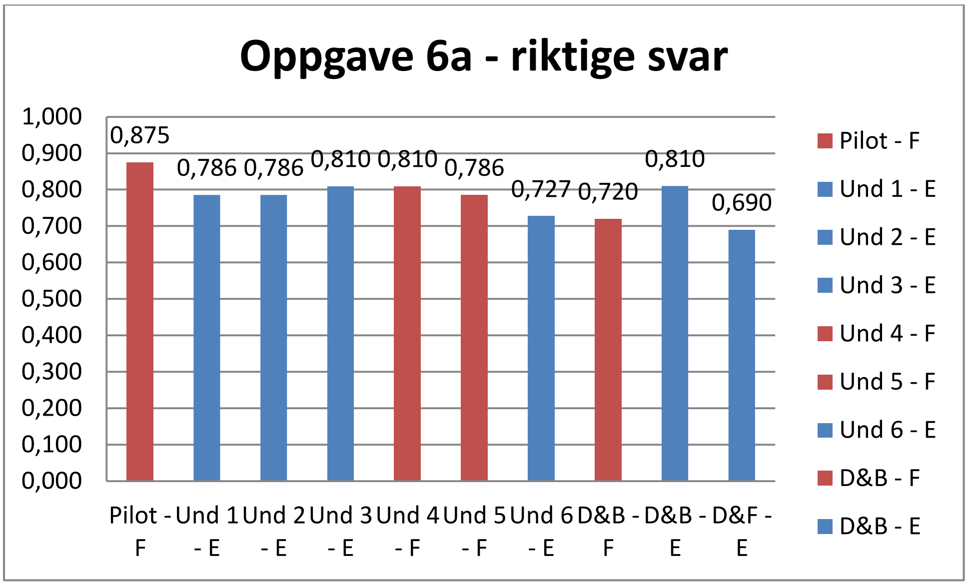
Det er to ulike deloppgaver i oppgave 6. Begge oppgavene er betinget sannsynlighet i et forsøk uten tilbakelegging, men for oppgave 6b har vi omvendt tidsakse, beskrevet under misoppfatning 3B, inversjon av tidsaksen.

Oppgave 6
To svarte og to hvite kuler ligger i en urne. Vi trekker ut to kuler etter hverandre fra urnen uten å legge dem tilbake. *(P betyr sannsynlighet, W betyr hvit og nummer er hvilken kule)*

a) Dersom den første kulen er hvit, hvor stor er sannsynligheten for at den andre kulen er hvit?
$P(W_2 | W_1)$: i) ½ ii) 1/6 **iii) 1/3** iv) ¼

b) Dersom den andre kulen er hvit, hvor stor er sannsynligheten for at den første kulen er hvit?
$P(W_1 | W_2)$: **i) 1/3** ii) umulig å avgjøre iii) 1/6 iv) ½

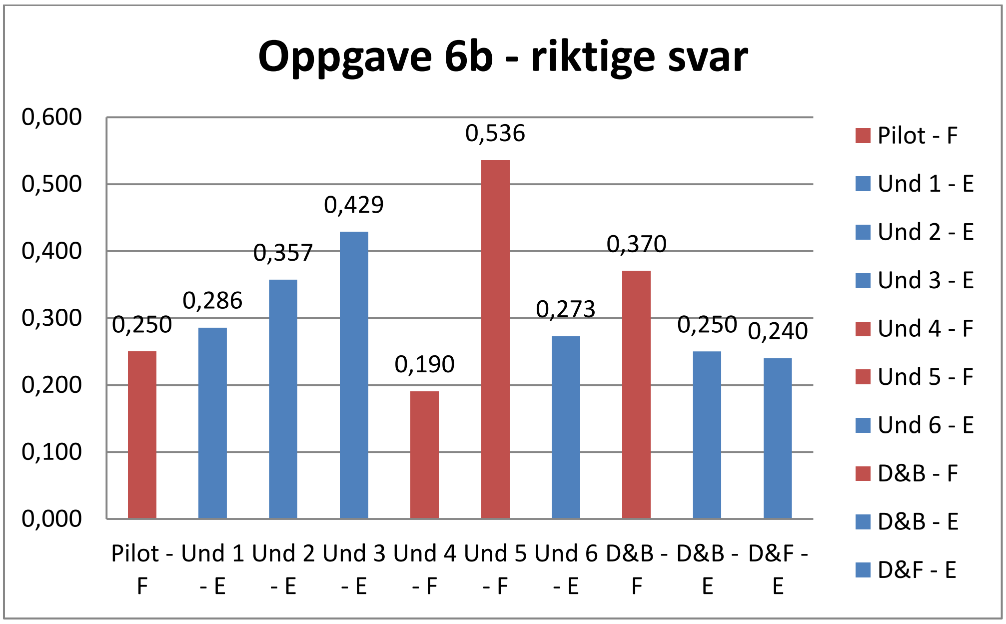
På grunn av forskjellen mellom oppgave 6a og 6b blir resultatene presentert hver for seg. Vi vil også se at det er stor forskjell på prestasjonene mellom oppgave 6a og oppgave 6b.



Figur: Oppgave 6a. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Vi ser at det er liten variasjon mellom gruppene. Alle presterer høyt med 0,69 som lavest (med undervisning) og 0,875 som høyest (uten undervisning). Basert på resultatene i de ulike gruppene til oppgave 6a, er det vanskelig, om ikke umulig, å si at undervisning har noen nytte. Igjen seiler oppgaver som er kompatible med tidsaksen opp som rent intuitive og som de fleste forstår. Derfor er det mange forskere som taler varmt for at denne typen oppgaver, som henger sterkt sammen med valgtre, kan komme på et tidligere tidspunkt i undervisningen. Det er tydelig ingen særlige sterke misoppfatninger som hindrer forståelse av oppgaver med suksessive forsøk som er kompatible med tidsaksen.

For oppgave 6b, med omvendt tidsakse, er derimot situasjonen annerledes. Diagrammet viser andelen riktige svar for hver gruppe.



Figur: Oppgave 6b. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Den ene gruppe som har har en score over 50 prosent har ikke hatt undervisning i emnet. En annen interssant observasjon er at blant de gruppene som har hatt undervisning, scorer begge S1-gruppene (Und2 og Und3), altså på vg2-nivå, bedre enn gruppene på vg1-nivå (Und1 og Und6). Ut i fra presumtivt bedre elever på vg1 og resultater fra de andre oppgavene i testen, skiller altså denne oppgaven seg ut. Elevene på vg2 har en betraktelig høyere prestasjon i akkurat i denne oppgaven. Mest bemerkelsverdig er den store nedgangen for Diaz med flere, samt nedgangen for Sam2 (Und5 og Und6). Det kan se ut som om at undervisning snarere forsterker misoppfatningen om at en årsak må komme før effekten (the Falk phenomenon), enn å åpne opp for at så kan skje. Vi kan se samme tendens som i oppgave 5; misoppfatningen forsterkes med undervisning. De fire beste gruppene er de litt svakere S1-gruppene og to 1T-gruppene før undervisning. Se drøfting i kapittel 5.

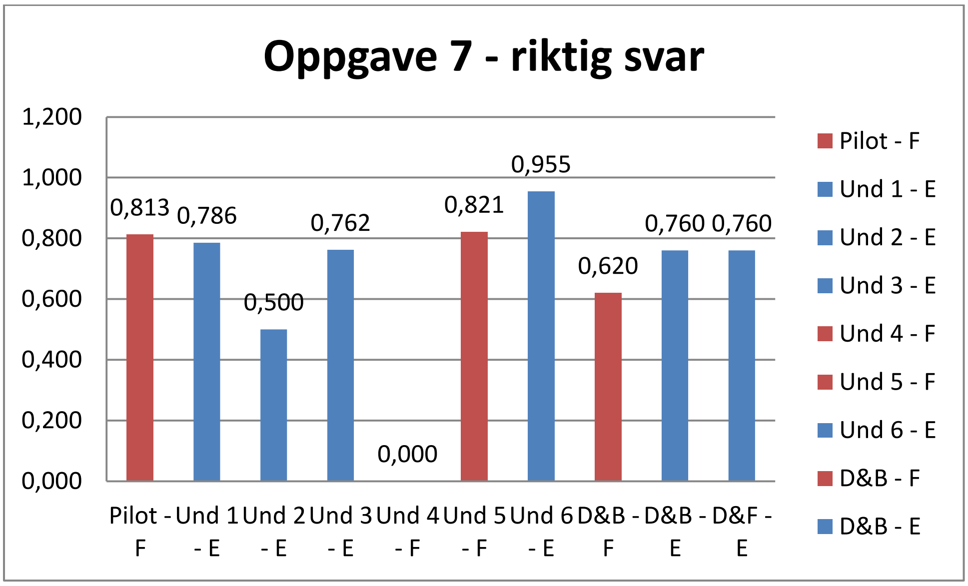
4.7 Resultat oppgave 7

Oppgave 7 ligner på oppgave 2, men er noe mer avansert, fordi vi skal beregne sannsynlighet med produktsetningen i en asynkron situasjon. Vi har to avhengige hendinger. Vi kjenner igjen konteksten fra norsk skole, men som tidligere nevnt er sammenligning av sannsynligheter lite brukt.

Oppgave 7
En urne inneholder en blå og to røde kuler. Vi trekker ut to kuler, en etter en, uten å legge dem tilbake. Hvilken av hendingene nedenfor er mest sannsynlig eller er de like sannsynlige?

a) Få to røde kuler
b) Den første kulen er rød og den andre kulen er blå.
c) **De to hendingene a og b er like sannsynlige.**

Dersom elevene oppdager at vi først trekker en rød kule uansett, er det stor sjanse for å svare riktig. Und4 tok ikke denne oppgaven.



Figur: Oppgave 7. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Klart beste prestasjon er Und6 med nesten 96 prosent uttelling. Dette er en gruppe som har hatt undervisning Alle gruppene presterer på 50 prosent eller bedre. Jeg vil påstå at undervisningen har hjulpet etter direkte sammenligning mellom Und5-Und6 (Sam2) og mellom gruppene til Diaz med flere, selv om Sam1 ikke viser en forbedring. Når effekt følger årsak presterer elevene bedre.

4.8 Resultat oppgave 8

Denne oppgaven er den første av de åpne oppgavene. I dette kapitlet vil jeg ikke komme nærmere inn på delvis riktige svar, det blir drøftet i kapittel 5. Dette gjelder alle de åpne oppgavene. I denne oppgaven er hovedmomentet at elevene gjenkjenner at en betingelse begrenser utfallsrommet. Det er altså ikke en egen misoppfatning, men essensen i det hele. Misoppfatning 7, kombinatoriske problem, kan være et hinder på prestasjonssiden. Dette drøftes i kapittel 5.
Oppgave 8 gir svar på evnen til å finne det fullstendige utfallsrommet, som er ett av fire viktige steg til å forstå sannsynlighet (side 12 i Bryant og Nunes, 2012). De som ikke lister alle mulige rekkefølger har en kombinatorisk misoppfatning. Denne kombinatoriske misoppfatningen er et hinder i å utarbeide et fullstendig utfallsrom og dermed en korrekt sannsynlighet (som det ikke var spurt etter denne gangen).

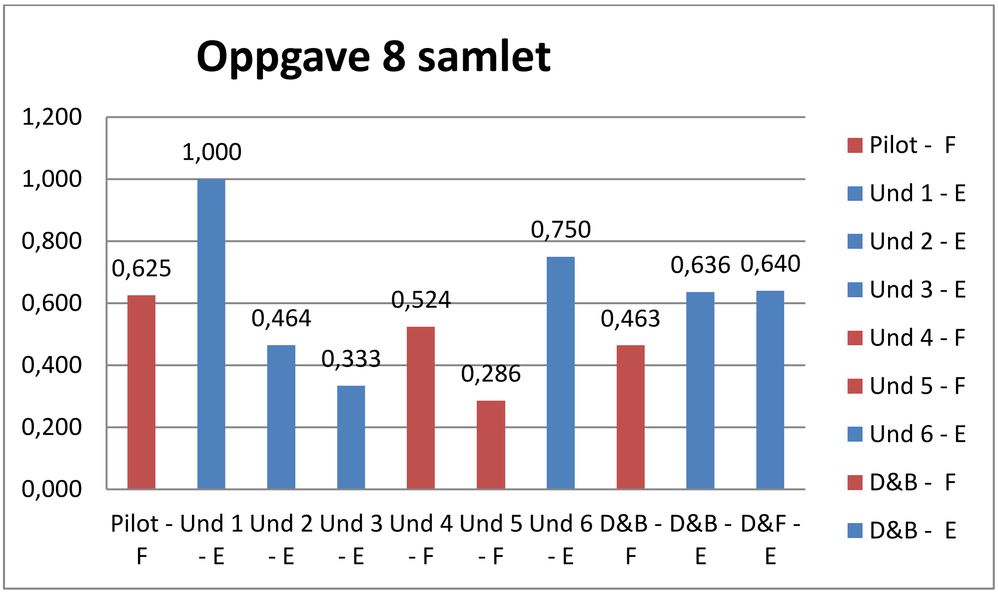
Oppgave 8
Lag utfallsrommet ferdig i de følgende tilfeldige eksperimenter:

a) Kjønnsfordelingen (Gutt/Jente) til tre barn i en familie (For eksempel GJG,…)
b) Kjønnsfordelingen (Gutt/Jente) til tre barn i en familie dersom det er to eller flere gutter.

De riktige svarene er:

a) $U = \{GGG,\ GGJ,\ GJG,\ JGG,\ GJJ,\ JGJ,\ JJG,\ JJJ\}$
b) $U_b = \{GGG,\ GGJ,\ GJG,\ JGG\}$

På grunn av Batanero og Diaz presenterer oppgave 8a og 8b samlet som oppgave 8, vil jeg gjøre det samme. I min undersøkelse har jeg da tatt gjennomsnittet av resultatene i oppgave a og b. Grundig analyse (i kapittel 5) viser at denne verdien er tilnærmet riktig. Resultatet er som følger:



Figur: Oppgave 8 Samlet. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Hos Diaz med flere ser vi en økning i prestasjon fra gruppen uten undervisning til gruppene med undervisning. I denne oppgaven har studentene til Batanero og Diaz en forbedring i antall riktige svar på 17,3 prosentpoeng, fra 46,3 % til 63,6 %. I mine undersøkelser viser Sam1 og Sam2 en betydelig forbedring i resultat. Prestasjonen til Und1 er imponerende og kan ikke skyldes annet enn undervisning som har vektlagt utarbeidelse av komplette utfallsrom. En undervisning som hjelper elevene med å løse kombinatoriske floker er essensielt for å svare riktig på slike oppgaver (side 211 i Polaki, 2005). Oppgave 8 viser at undervisning understøtter elevenes evne til å begrense utfallsrommet etter at betingelser er gitt. Allerede i den første undersøkelsen til Diaz og Fuente viste mange studenter at de behersket innsnevring av utfallsrommet. Hele 85 prosent av studentene svarte riktig eller delvis riktig på denne oppgaven (Diaz og Fuente, 2007).

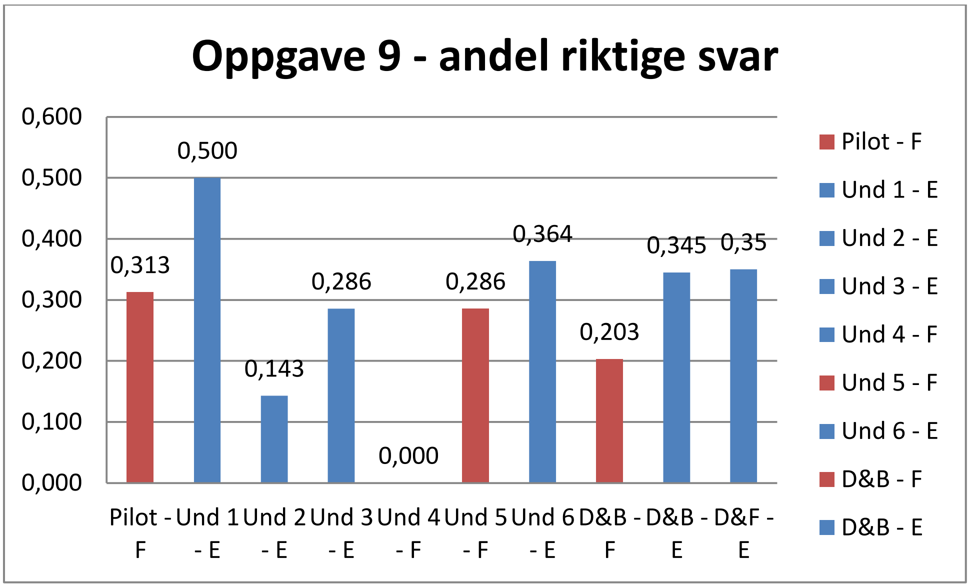
4.9 Resultat oppgave 9

I oppgave 9 skal elevene beregne betinget sannsynlighet i ett forsøk. Oppgaven har likhetstrekk med oppgave 8 ved at utfallsrommet må begrenses. Noen feilaktige svar har bakgrunn i heuristikker. Dette drøftes i kapittel 5. I dette kapitlet vurderes det om beregningene er riktige.

Oppgave 9
Ved å trille to terninger (en grønn og en rød) får vi at produktet av de to tallene som terningene viser er lik 12. Hvor stor er sannsynligheten for at ingen av de to terningene er en sekser?

Elevene må ut fra den gitte betingelsen begrense utfallsrommet fra 36 til fire utfall. De mulige utfallene er {(2,6), (3,4), (4,3), (6,2)}. De gunstige utfallene er {(3,4), (4,3)}, og ved å anvende Laplace-regelen får vi svaret 0,5. Merk at svaret også blir riktig uten komplett kombinatorikk: (3,4) og (2,6) gir (3,4) og sannsynligheten 0,5.

Av diagrammet nedenfor ser vi at Und4 ikke hadde denne oppgaven.



Figur: Oppgave 9. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

I denne oppgaven har studentene til Diaz og Batanero en forbedring i antall riktige svar på 14,2 prosentpoeng, fra 20,3 % til 34,5 % (Diaz og Batanero, 2009). Direkte sammenligning med Sam1 og Sam2 viser i gjennomsnitt en tilsvarende økning med om lag 19 prosentpoeng (Sam1) og 8 prosentpoeng (Sam2), gjennomsnittlig omtrent som Diaz og Batanero. Vi ser at S1-gruppene (Und2 og Und3) generelt scorer lavt. Som for oppgave 3, 4 og 5 er alle prestasjoner på 50 prosent eller lavere. Dette er ikke imponerende på noen måter, men undervisning hjelper elevene til å oppdage at betingelser gir begrensninger på utfallsrommet.

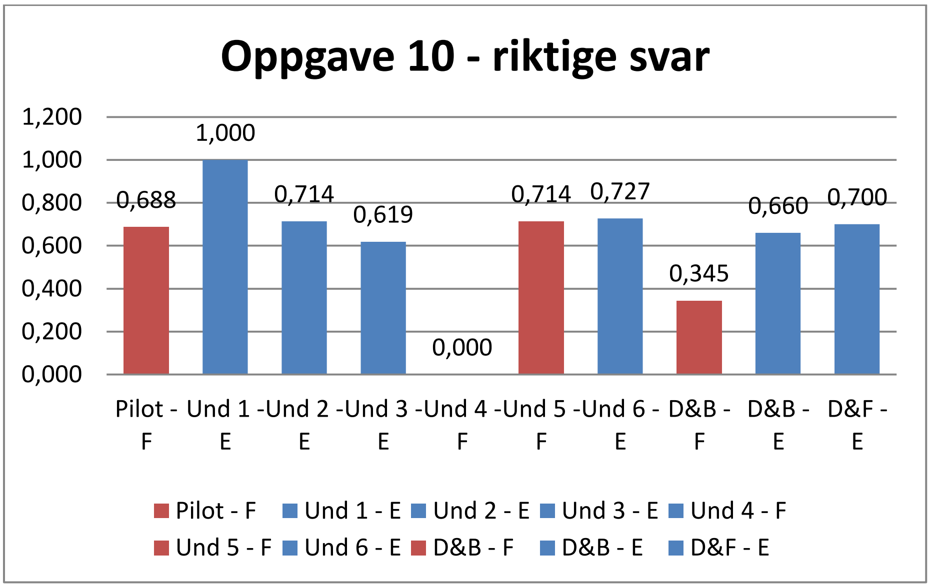
4.10 Resultat oppgave 10

Elevene skal beregne «betinget» sannsynlighet i et forsøk med tilbakelegging, altså uavhengighet. Å vite om vi har et forsøk med eller uten betingelser, eller med eller uten tilbakelegging, er en del av kunnskapen knyttet til betinget sannsynlighet. Heuristikker knyttet til denne oppgaven drøftes i kapittel 5.

Oppgave 10
En person kaster en rettferdig terning og skriver ned resultatet som oddetall eller partall. Han får dette resultatet etter 15 kast: o p p o o p o o o o p p o o o
Personen kaster terningen en gang til. Hvor stor er sannsynligheten for å få et oddetall nå?

For å løse denne oppgaven må eleven forstå at tidligere forsøk ikke påvirker utfallene til kommende forsøk, siden hvert terningkast er uavhengige av hverandre. Ordet uavhengig er ikke presisert i oppgaven, men terningen er rettferdig og sannsynligheten er lik 0,5 for oddetall ved neste kast.

Av diagrammet nedenfor ser vi at Und4 ikke hadde denne oppgaven.



Figur: Oppgave 10. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

Her ser vi noenlunde samme tendens som i oppgave 8, men med høyere score. Over 50 prosent av elevene/studentene i hver gruppe svarer riktig, unntatt er D & B-F (Batanero og Diaz, før undervisning). Studentene til Batanero og Diaz øker andelen riktige svar med 31,5 prosentpoeng, fra 34,5 % til 66,0 %. I mine undersøkelser viser Sam1 en tilsvarende økning, mens Sam2 har en knapp økning. De to S1-gruppene presterer noenlunde likt – og bra.

Ut fra dette vil jeg si at undervisning hjelper med å løse problemer med betinget sannsynlighet for uavhengige suksessive hendinger, eller å belyse når vi har uavhengige hendinger. Som i oppgave 8 er prestasjonen til Und1 imponerende, og resultatet må tilskrives en undervisning som peker på forskjellen mellom avhengighet og uavhengighet. Min erfaring forteller meg også at et grundig arbeid med valgtrær hjelper elevene til å se forskjell på hendinger med og uten tilbakelegg.

4.11 Resultat oppgave 11

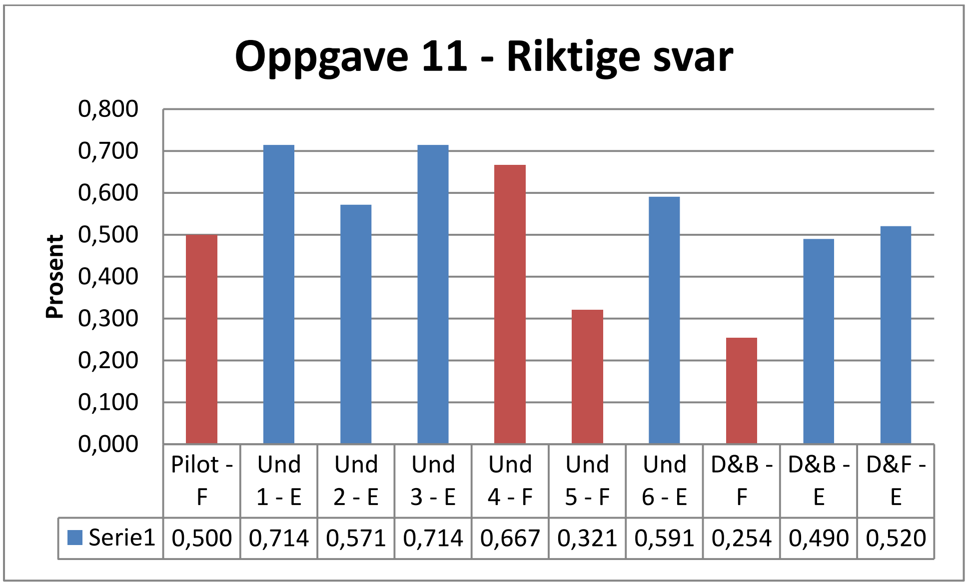
Som i oppgave 10 har vi også her uavhengige hendinger, men her må elevene trekke inn produktsetningen. Jamfør oppgave 2 og 7 for avhengige hendinger. Konteksten er etter forslag fra veileder endret slik at de to hendingene i større grad har et preg av uavhengighet. Ordet uavhengig var også med i den opprinnelige teksten.

Oppgave 11
I en gruppe elever på en skole svarer 80 % av elevene at de har øvelseskjørt med foreldrene 10 ganger eller mer uten å kollidere. Av de samme elevene svarer 70 % at de bestod den siste testen i engelsk. Anta at hva en elev klarer, å bestå en test eller å øvelseskjøre uten kollisjon, er uavhengige av hverandre.
Hvor stor er da sannsynligheten for at en tilfeldig valgt elev har bestått engelsktesten og øvelseskjørt mer enn 10 ganger uten å kollidere?

For å løse dette problemet må elevene anvende produktsetningen for uavhengige hendinger. La K være hendingen feilfri kjøring og E være hendingen bestått engelsktesten:

$$P(K \cap E) = P(K) \cdot P(E) = 0,80 \cdot 0,70 = 0,56$$

Riktig symbolbruk eller den lille gangetabellen er ikke nødvendig for at svaret skal betraktes som riktig. De to faktorene i mellomregningen er for eksempel godt nok.



Figur: Oppgave 11. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

I denne oppgaven har studentene til Batanero og Diaz en forbedring i antall riktige svar på 23,6 prosentpoeng, fra 25,4 % til 49,0 %. I den første undersøkelsen til Diaz og Fuente viste mange studenter at de ikke behersket sammensatte og uavhengige hendinger. Kun 52 % svarte riktig på denne oppgaven, mens hele 30 % svarte blankt eller helt feil (Diaz og Fuente, 2007). Dersom vi ser på sammenhengen mellom Pilot og Und1 (Sam1) og mellom Und5 og Und6 (Sam2), viser dette samme tendens som Diaz og Batanero sine grupper før og etter undervisning. Det er en forbedring på henholdsvis 21 % (Sam1)og 27 % (Sam2) mot 25 % for D & B. Vi ser også at Und2 og Und3 har et bedre resultat enn pilot, mens Und4, uten undervisning, har gjort det bemerkelsesverdig bra. S1-gruppene Und2 og Und3 scorer bedre enn den sterke pilotgruppen. S1-gruppene har hatt undervisning i emnet på vg1 og vg2 før testen. Det indikerer at undervisning hjelper å oppdage uavhengighet mellom to hendinger, og riktig bruk av produktregelen.

4.12 Resultat oppgave 12

I denne oppgaven skal elevene beregne betinget sannsynlighet med produktsetningen i en synkron situasjon på to avhengige hendinger. Oppgaven tilsvarer oppgave 11, men med avhengige hendinger.

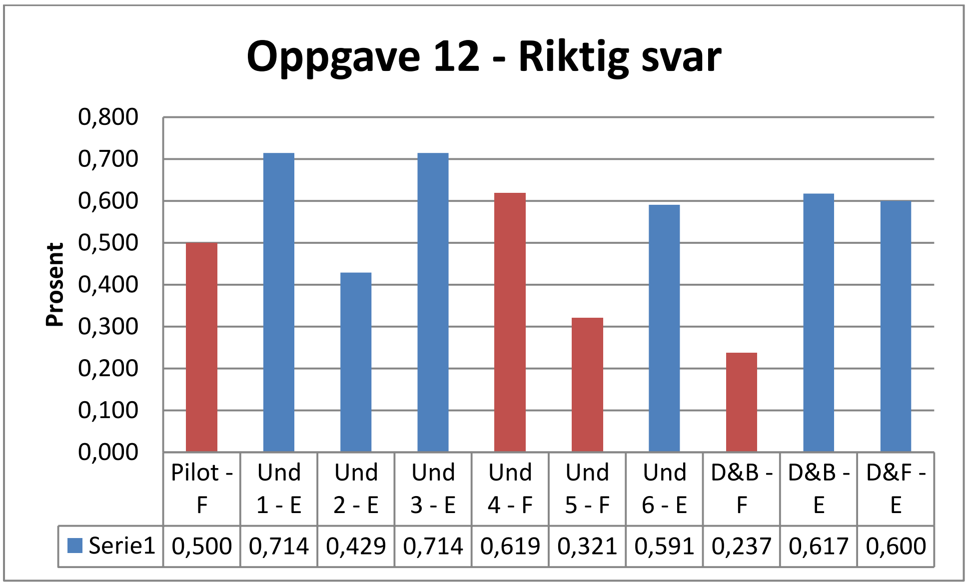
Oppgave 12
Ifølge nylig forskning så lyver 91 % av folkene i en populasjon (for eksempel Bergen) og 36 % av disse lyver om viktige saker. Dersom vi plukker en tilfeldig person fra byen, hvor stor er sannsynligheten for at denne personen lyver om viktige saker?

Dette problemet undersøker elevenes evne til å anvende produktregelen for to avhengige hendinger. La L være hendingen at en person lyver og V være hendingen at personen lyver om viktige saker:

$$P(L \cap V) = P(L) \cdot P(V \mid L) = 0{,}91 \cdot 0{,}36 = 0{,}327$$

Symbolbruk er ikke nødvendig. De to faktorene i mellomregningen er svar godt nok.

Diagrammet viser kun andel riktige svar.



Figur: Oppgave 12. Prosentandel riktige svar for hver gruppe. F = Før og E = Etter.

I den første undersøkelsen til Diaz og Fuente viste mange studenter at de behersket å løse problemer med i avhengige hendinger. Hele 60 % svarte riktig på denne oppgaven (Diaz og Fuente, 2007). Det som først og fremst er bemerkelsesverdig er den store likheten i resultatene mellom denne oppgaven og oppgave 11. Faktisk er halvparten av andelen riktige svar i oppgave 11 og 12 identiske, for eksempel Und3 som presterer 71,4 % på begge oppgavene. Studentene til Batanero og Diaz forbedrer resultatet med 38,0 prosentpoeng, fra 23,7 % til 61,7 %. Fra Pilot til Und1 (Sam1) er økningen på vel 21 %. Fra Und5 til Und6 (Sam2) er økningen 27 %. Vi ser også at alle de blå stolpene er høyere enn de to laveste røde stolper, og at de to beste er blå. På oppgave 11 var det enda litt bedre. Selv om oppgaven i CTR er beskrevet med en synkron situasjon, kan elevene lett tenke seg dette som en asynkron situasjon, og hendingene følger etter hverandre på tidsaksen. Bruk av produktsetningen når hendingene er kompatible med tidsaksen fungerer for elevene, selv om rekkefølgen av hendingene fremstår som uklar. Det kommer klart frem at undervisning hjelper.

4.13 Resultat oppgave 13

Denne oppgaven kommer litt på siden av de andre. Vi ser etter forbedringer i prestasjonene i å løse et bayesisk problem. Oppgaven er nøye drøftet i CPR, men på grunn av elevenes manglende forkunnskap til bayesiske problemer og den lave prestasjonen i mine undersøkelser, vil jeg ikke drøfte oppgaven i kapittel 5. Jeg får lite ut av denne oppgaven, bortsett fra at svarene uteblir, noe som er et svar i seg selv.

Oppgave 13
To maskiner, M1 og M2, produserer baller. Maskin M1 produserer 40 % og maskin M2 produserer 60 % av ballene. Av ballene som blir produsert av maskin M1 er det 5 % som er defekte og av de som blir produsert av maskin M2 er det 1 % som er defekte. Vi plukker ut en tilfeldig ball og finner ut at den er defekt. Hvor stor er sannsynligheten for at ballen var produsert i maskin M1?

Det er flere mellomregninger som skal gjøres, og det er derfor strengt tatt mer problematisk å komme frem til riktig løsning. La M1 og M2 være hendingene som viser produksjonsmaskin og D være hendingen at en ball er defekt:

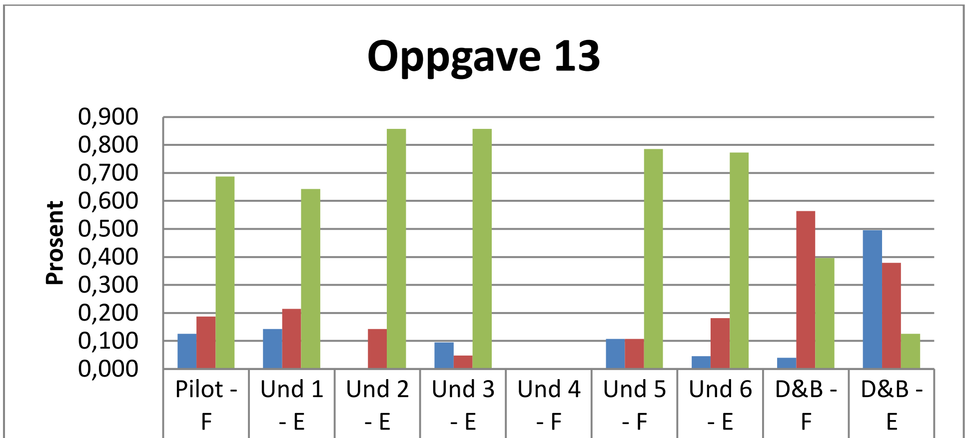
$$P(M_1 \mid D) = \frac{P(D \cap M_1)}{P(D)} = \frac{P(M_1)\cdot P(D \mid M_1)}{P(M_1)\cdot P(D \mid M_1) + P(M_2)\cdot P(D \mid M_2)}$$

$$= \frac{0,40\cdot 0,05}{0,40\cdot 0,05 + 0,60\cdot 0,01} = \frac{0,02}{0,026} = 0,769$$

Forfatterne Batanero og Diaz identifiserer fire punkter for å få et riktig svar.

1. Identifisere riktige data
2. Identifisere og beregne den betingede sannsynligheten i telleren
3. Beregne riktig nevner, total sannsynlighet
4. Beregne den omvendte sannsynligheten (Bayes' setning)

Alle alternativer for svar tatt med i stolpediagrammet nedenfor. Vi merker at gruppen til Diaz og Fuente ikke er med. Jeg fant ikke tall til denne oppgaven. Gruppen Und4 hadde ikke denne oppgaven. Diagrammet viser andel riktige, delvis riktige og feilaktige svar.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E	D&B - F	D&B - E
Riktig	0,125	0,143	0,000	0,095	0,000	0,107	0,045	0,040	0,495
Delvis	0,188	0,214	0,143	0,048	0,000	0,107	0,182	0,564	0,379
Feil	0,688	0,643	0,857	0,857	0,000	0,786	0,773	0,396	0,126

Figur: Oppgave 13. Prosentandel for riktige, delvis riktige ogfeilaktige svar for hver gruppe.

For elevene i mine undersøkelser er det liten eller ingen bedring å spore. Mine resultater samsvarer ikke Diaz og Batanero. Konklusjonen er at jeg ikke kan se noen sammenhenger som viser endring i resultat mellom før og etter undervisning for mine grupper. Resultatet er like svakt uansett. Dette harmonerer med at ingen spesiell undervisning er gitt i bruk av Bayes' setning. Gruppen D & B – E har hatt undervisning i bayesiske problemer og dette kan forklare den markante forbedringen; fra 4 % til 50 % andel riktige svar. Diaz og Batanero kan konkludere med at undervisning hjelper, men mine resultater kan ikke støtte dette. Et interessant spørsmål er hvorvidt resultatet ville sett annerledes ut dersom oppgaven var presentert med frekvenser. Forskning viser at resultatet da blir bedre (Side 71 i Watson og Moritz, 2002).

4.14 Resultat fra CPR

Jeg har materiale fra tre undersøkelser fra Diaz med flere. Jeg vil i dette avsnittet stort sett referere til de to siste undersøkelsene (Diaz og Batanero, 2009), men noen ganger vil jeg ta med den første undersøkelsen også (Diaz og Fuente, 2007). Tabellen nedenfor viser det samlede resultat fra de to siste undersøkelsene. Dette resultatet er basert på følgende (Diaz og Batanero):

- Max poeng er 30
- 14 multiple choice-oppgaver som hver gir 1 poeng
- 8 åpne svar oppgaver som gir 2 poeng for korrekt svar og 1 poeng for delvis riktig svar

Gruppe	n = antall	Gjennomsnitt	Standardavvik	Standardfeil
D & B – F (Før undervisning)	177	12,68	5,69	0,429
D & B – E (Etter undervisning)	206	18,47	5,15	0,358

Tabell: Samlet resultat. Batanero og Diaz. Grupper uten og med undervisning.

Resultatet indikerer en markant forbedring i formell forståelse av betinget sannsynlighet og i evnen til problemløsing, men liten endring i oppgaver som er forbundet med heuristikker (psychological biases). Forskerne Batanero og Diaz konkluderer med en statistisk signifikant forskjell i favør D & B - E, studenter med undervisningsøkt i sannsynlighetsregning. Dette baserer de på en t-test med $t = 8,61$ (p<0,0001), som jeg ikke går videre inn på. Vi kan observere at gjennomsnittspoengsummen øker fra 12,68 poeng til 18,47 poeng. Diaz og Batanero argumenterer videre med at nedre og øvre kvartil, minimum og maksimum og median, alle lå høyere for gruppe «Etter» enn for «Før». Sentralmålene er altså forskjøvet mellom gruppene, men spredningsmålene som standardavvik og boks-plot, er noenlunde identiske. Dette er et klart tegn på CPR-testens validitet, vel som konklusjonen om at undervisningsøkten har hjulpet. En analyse av resultatene til CPR-testen viste at de med 82,3 % sikkerhet kunne fastslå om en student tilhørte «Før» eller «Etter». Det neste steget i undersøkelsen til Diaz og Batanero var å sjekke om forbedringen gjaldt på alle oppgavene i testen. Resultat fra oppgaver i CPR som ikke er med i denne undersøkelsen er utelatt og jeg forholder meg til min oppgavenummering. Tabellen nedenfor viser prosentandelen riktige svar for begge gruppene.

Oppg.	**Enhet**	**Beskrivelse av emne (enhet)**	**D & B – F**	**D & B – E**
1a	4	Beregne sannsynlighet for enkel hending fra krysstabell	35	69
1b	4 og 15	Beregne betinget sannsynlighet fra krysstabell	67	94
1c	4 og 15	Beregne sannsynlighet for snitt mellom to hendinger fra krysstabell	29	63
1d	4 og 15	Beregne omvendt (fra 1b) betinget sannsynlighet fra krysstabell	37	70
2	10 og 14	Løse et problem med betinget sannsynlighet ved avhengighet	77	89
3	6	Snitt i sammensatte hendinger. «Conjunction fallacy»	21	24
4	5 og 13	Omvendt betinget / årsak - diagnostisk	35	35
5	12 og 14	Omvendt tidsakse – beregne betinget sannsynlighet	8	13
6a	10 og 14	Beregne betinget sannsynlighet uten tilbakelegging (asynkron)	72	81
6b	10, 12 og 14	Som ovenfor, men med omvendt tidsakse (enhet 12)	37	25
7	14	Løse et problem med betinget sannsynlighet i et asynkront forsøk	62	76
8	2	Gjenkjenne at en betingelse gir en begrensning av utfallsrommet	46	64
9	8	Beregne betinget sannsynlighet i ett enkelt forsøk	20	35
10	9	Beregne betinget sannsynlighet i et forsøk med tilbakelegging	35	60
11	16	Bruke produktsetningen på to uavhengige hendinger	26	49
12	17	Bruke produktsetningen på to avhengige hendinger	24	62
13	19	Løse bayesisk problem	4	50

Tabell: Prestasjon på hver oppgave CPR. Mine oppgavenummer. Enhet fra CPR. Prosentscore.

Oppgaver med grønn farge viser markant forbedring etter undervisningsøkten, mens rød farge viser en forverring. Oppgave 6b er den eneste oppgaven som viser en forverring etter undervisningsøkten, mens oppgave 4 viser ingen forskjell. Oppgave 3 og 5 viser liten endring i favør «Etter». Alle de andre oppgavene viser en forbedring av prestasjonene i favør «Etter», og er ifølge Batanero og Diaz statistisk signifikant. Vi merker oss to oppgaver med omvendt tidsakse, oppgave 5 og 6b. Dersom vi ser bort fra psykologiske feil til studenter som nekte å reversere betingelser til tidligere hendinger, kan CPR-testen brukes til å skille mellom studenter med og uten spesiell undervisningsøkt.
Alle åpne oppgaver har en forbedring i prestasjonen for gruppen med undervisning.
Batanero og Diaz påstår altså at undervisning reduserer misoppfatninger innen sannsynlighet (unntatt feilslutningen om den omvendte tidsakse, enhet 12).

4.15 Samlet resultat

Vil mine kvantitative undersøkelser gi de samme svarene? Jeg søker svar på:

- **Hvilken innvirkning har undervisning i skolen i Norge innen betinget sannsynlighet?**
- **Vil elever redusere antall feilslutninger etter gjennomgått undervisning?**

Tabellen nedenfor har grå felter for de oppgavene som ikke var med i undersøkelse 4. De blå feltene på oppgave 8 for Diaz med flere har sammenheng med at oppgave 8a og 8b ble slått sammen (Diaz og Fuente, 2007/Diaz og Batanero, 2009). De grønne og oransje feltene øverst i hver kolonne markerer grupper før og etter undervisning.

Oppg	pilot - F	Und1 - E	und2 - E	und3 – E	und4 - F	und5 - F	Und6 - E	D & B - F	D & B - E	D & F - E
1a	87,5	100,0	85,7	100,0	85,7	85,7	90,9	35,0	69,0	90,0
1b	81,3	57,1	57,1	61,9	52,4	60,7	72,7	67,0	94,0	61,0
1c	93,8	92,9	71,4	90,5	52,4	71,4	81,8	29,0	63,0	59,0
1d	75,0	92,9	57,1	61,9	52,4	71,4	86,4	37,0	70,0	56,0
2	81,3	85,7	71,4	85,7	0,0	85,7	90,9	77,0	89,0	89,0
3	37,5	50,0	28,6	19,0	0,0	17,9	27,3	21,0	24,0	25,0
4	50,0	35,7	21,4	23,8	0,0	21,4	22,7	35,0	35,0	32,0
5	18,8	7,1	35,7	19,0	28,6	17,9	13,6	8,0	13,0	10,0
6a	87,5	78,6	78,6	81,0	81,0	78,6	72,7	72,0	81,0	69,0
6b	25,0	28,6	35,7	42,9	19,0	53,6	27,3	37,0	25,0	24,0
7	81,3	78,6	50,0	76,2	0,0	82,1	95,5	62,0	76,0	76,0
8a	56,3	100,0	35,7	28,6	57,1	25,0	68,2	46,3	63,6	64,0
8b	68,8	100,0	57,1	38,1	47,6	32,1	81,8	0,0	0,0	
9	31,3	50,0	14,3	28,6	0,0	28,6	36,4	20,3	34,5	35,0
10	68,8	100,0	71,4	61,9	0,0	71,4	72,7	34,5	66,0	70,0
11	50,0	71,4	57,1	71,4	66,7	32,1	59,1	25,4	49,0	52,0
12	50,0	71,4	42,9	71,4	61,9	32,1	59,1	23,7	61,7	60,0
13	12,5	14,3	0,0	9,5	0,0	10,7	4,5	4,0	49,5	

Tabell. Samlet resultat i prosent fra alle undersøkelser.

Resultatet i tabellen bekrefter hypotesen om at undervisning hjelper med formell forståelse av betinget sannsynlighet. Dette blir tydeligere i de neste tabellene, men jeg vil likevel diskutere noen tall fra mine undersøkelser:

- På oppgave 8a ser vi at S1-gruppene har størst problemer med å utarbeide komplette utfallsrom (gul).
- På oppgave 8 og 10 ser vi at Und1 presterer 100 prosent, langt bedre enn de andre gruppene (grønn).
- På oppgave 11 og 12 ser vi at Und5 presterer langt dårligere enn de andre gruppene (rød).

For den neste tabellen trengs det litt informasjon. Fargekodene og kolonnene er viktige.

- Direkte sammenligning: Før = Pilot og Und5. Etter = Und1 og Und6 (altså et utvalg)
- Batanero og Diaz: Ikke med resultat fra Diaz og Fuente (opprettholde balanse)
- Oransje: Merk at Batanero og Diaz har slått sammen resultat på oppgave 8.
- Alle grupper: Alle mine undersøkelser fordelt i *før* og *etter* undervisning.
- Grå felt: Oppgaver der Und4 ikke svarte. Grunnlaget er da 44 i stedet for 65.
- Røde oppgaver i første kolonne: Negativ trend etter undervisning
- Grønne oppgaver i første kolonne: Positiv trend etter undervisning
- Røde prosenter: Negativt resultat. Score etter undervisning er lavere enn før undervisning.

Fargeleggingen av trenden i oppgavene er min egen subjektive oppfatning basert på mine resultater, og er ikke relatert til noe tallmateriale fra Batanero og Diaz.

	Direkte sammenligning			Alle grupper			D & B		
Antall	n = 44	n = 36		65	71	44	n=177	n=206	
Oppgave	**Før**	**Etter**	**Endring**	**Før**	**Etter**	**Endring**	**Før**	**Etter**	**Endring**
1a	0,864	0,944	8,1 %	0,862	0,944	8,2 %	0,350	0,690	34,0 %
1b	0,682	0,667	-1,5 %	0,631	0,634	0,3 %	0,670	0,940	27,0 %
1c	0,795	0,861	6,6 %	0,708	0,845	13,7 %	0,290	0,630	34,0 %
1d	0,727	0,889	16,2 %	0,662	0,746	8,5 %	0,370	0,700	33,0 %
2	0,841	0,889	4,8 %	0,569	0,845	27,6 %	0,770	0,890	12,0 %
3	0,250	0,361	11,1 %	0,250	0,296	4,6 %	0,210	0,240	3,0 %
4	0,318	0,278	-4,0 %	0,318	0,254	-6,5 %	0,350	0,350	0,0 %
5	0,182	0,111	-7,1 %	0,318	0,183	-13,5 %	0,080	0,130	5,0 %
6a	0,818	0,750	-6,8 %	0,815	0,775	-4,1 %	0,720	0,810	9,0 %
6b	0,432	0,278	-15,4 %	0,354	0,338	-1,6 %	0,370	0,250	-12,0 %
7	0,818	0,889	7,1 %	0,818	0,775	-4,4 %	0,620	0,760	14,0 %
8a	0,364	0,806	44,2 %	0,431	0,563	13,3 %	0,463	0,636	17,3 %
8b	0,455	0,889	43,4 %	0,462	0,676	21,5 %	0,000	0,000	0,0 %
9	0,295	0,417	12,1 %	0,295	0,324	2,8 %	0,203	0,345	14,2 %
10	0,705	0,833	12,9 %	0,705	0,746	4,2 %	0,345	0,660	31,5 %
11	0,386	0,639	25,3 %	0,477	0,648	17,1 %	0,254	0,490	23,6 %
12	0,386	0,639	25,3 %	0,462	0,620	15,8 %	0,237	0,617	38,0 %
13	0,114	0,083	-3,0 %	0,114	0,070	-4,3 %	0,040	0,495	45,5 %

Tabell: Samlet resultat alle oppgaver for tre grupperinger. Prestasjon før, prestasjon etter og endring i prosent.

Ut i fra tabellen leser jeg at undervisning hjelper fordi elevene generelt sett presterer bedre etter undervisning. Tabellen viser prestasjonen på 18 oppgaver og i begge kolonnene (4 og 7) ser vi at det er et negativt resultat i seks oppgaver i hver kolonne. Det betyr at det er positivt resultat i 2/3 av oppgavene. Det virker kanskje ikke som et formidabelt resultat, men samtidig må vi observere at tallene til de positive endringene ofte er betraktelig høyere enn tilsvarende negative endringer. Generelt sier jeg derfor at undervisning hjelper.

På hvilke områder ble elevene målt? Tabellen nedenfor repeterer enhetene fra CPR-undersøkelsen, og den prosentvise endringen fra tabellen ovenfor. De fargelagte, men tomme områdene under endringer i prosent, er lagt inn fordi en oppgave kan omfatte flere enheter, som for eksempel oppgave 1c.

			Endringer i prosent		
Oppg	**nr**	**Enhet fra CPR**	Direkte	Alle	D & B
1a	4	Identifisere betingede, **enkle** og sammensatte sannsynligheter	8,1 %	8,2 %	34,0 %
1b	4	Identifisere betingede, enkle og **sammensatte** sannsynligheter	-1,5 %	0,3 %	27,0 %
1c	4	Identifisere **betingede,** enkle og sammensatte sannsynligheter	6,6 %	13,7 %	34,0 %
	15	Beregne betinget sannsynlighet i en synkron situasjon			
1d	4	Identifisere **betingede,** enkle og sammensatte sannsynligheter	16,2 %	8,5 %	33,0 %
	15	Beregne betinget sannsynlighet i en synkron situasjon			
2	10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging	4,8 %	27,6 %	12,0 %
	14	Beregne betinget sannsynlighet i en asynkron situasjon			
3	6	Feilslutningen om snitt i sammensatte hendinger (conjuction fallacy)	11,1 %	4,6 %	3,0 %
4	5	Identifisere en betinget sannsynlighet og den inverse sannsynligheten	-4,0 %	-6,5 %	0,0 %
	13	Identifisere betinget årsak- og diagnostisk situasjon			
5	12	Beregne betinget sannsynlighet i et forsøk med omvendt tidsakse	-7,1 %	-13,5 %	5,0 %
	14	Beregne betinget sannsynlighet i en asynkron situasjon			
6a	10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging	-6,8 %	-4,1 %	9,0 %
	14	Beregne betinget sannsynlighet i en asynkron situasjon			
6b	10	Beregne betinget sannsynlighet i et forsøk uten tilbakelegging	-15,4 %	-1,6 %	-12,0 %
	12	Beregne betinget sannsynlighet i et forsøk med omvendt tidsakse			
	14	Beregne betinget sannsynlighet i en asynkron situasjon			
7	14	Beregne betinget sannsynlighet i en asynkron situasjon	7,1 %	-4,4 %	14,0 %
	17	Beregne sannsynlighet med produktsetningen på to avhengige hendinger			
8a	2	Gjenkjenne at en betingelse gir en begrensning av utfallsrommet	44,2 %	13,3 %	17,3 %
8b			43,4 %	21,5 %	
9	8	Beregne betinget sannsynlighet i ett enkelt forsøk	12,1 %	2,8 %	14,2 %
10	9	Beregne betinget sannsynlighet i et forsøk med tilbakelegging	12,9 %	4,2 %	31,5 %
11	16	Beregne sannsynlighet med produktsetningen på to uavhengige hendinger	25,3 %	17,1 %	23,6 %
12	15	Beregne betinget sannsynlighet i en synkron situasjon	25,3 %	15,8 %	38,0 %
	17	Beregne sannsynlighet med produktsetningen på to avhengige hendinger			
13	19	Løse Bayesisk problem	-3,0 %	-4,3 %	45,5 %

Tabell: Samlet resultat. Endring i prosent fra før undervisning til etter undervisning for tre grupper.

Vi ser en sterkest positiv tendens i oppgave 1, 8, 11 og 12. Samtidig ser vi positive tendenser i oppgave 2, 9 og 10. Generelt kan vi si at elevene er blitt flinkere til å knekke tekst-koden. De er i stand til å lese og forstå teksten i en oppgave. Oppgave 1, 11 og 12 viser at de i større grad er i stand til å skille mellom de ulike sammensatte sannsynlighetene som kan forekomme i vanlige oppgavetekster. Oppgave 2, 8, 9 og 10 viser at elevene i større grad evner å tolke når teksten i oppgaven gir oss en begrensning av utfallsrommet eller ei. Oppgave 3 viser også en liten forbedring. Denne oppgaven blir omtalt spesielt senere.

De to oppgavene med sterkest negativ trend er oppgave 5 og 6b. Dette er to oppgaver med omvendt tidsakse. Jeg har tidligere uttalt at undervisning har en direkte negativ påvirkning til elevers tolkning av oppgaver med omvendt tidsakse. For oppgave 6b er det en sterk korrelasjon til mellom mine resultater og CPR-undersøkelsen. CPR har sitt eneste negative resultat med hele 12,0 prosent på denne oppgaven. Jeg har ikke tilsvarende korrelasjon mellom mine resultater og CPR på oppgave 5, men mine resultater viser en kraftig negativ trend på denne oppgaven. I mine undersøkelser ser vi også en negativ trend i å identifisere en betinget sannsynlighet og den inverse sannsynligheten i en årsak- og diagnostisk situasjon, slik vi finner den i oppgave 4.

I mange oppgaver er det vanskelig å konkludere eller å komme med uttalelser om prestasjonene til elevene. På overflaten ser resultatet bra ut, men er det gjerne ikke. Det kan for eksempel være oppgaver som allerede har høy score **før** undervisningen starter. Fra tabellen nevner jeg oppgave 1a, 2, 6a og 7 som alle har en score på over 80 prosent før undervisning. Til sammenligning finner vi den høyeste scoren på CPR i oppgave 2 med 77 prosent før undervisning. Hvordan skal vi forbedre det? Oppgaver med høy score før undervisning har mindre sjanse for å måle en positiv tendens. Motsatt har oppgaver med lav score før undervisning større sjanse for å måle en positiv tendens. I mange oppgaver har CPR et lavt resultat før undervisning, noe som reder grunnen for et bedre resultat etter undervisning. I mine undersøkelser har vi en omvendt situasjon, fordi elevene også scorer høyt før undervisning. Sagt på en annen måte og i skolesammenheng: Det er betraktelig lettere for en elev som har 1 i standpunktkarakter å forbedre karakteren sin på en eksamen enn for en annen elev som har 6 i standpunktkarakter. Dette er delvis årsaken at jeg ikke vil plassere oppgave 6a under kategorien «undervisning nytter ikke». Oppgave 6a måler enhetene 10 og 14 fra tabellen:

10 Beregne betinget sannsynlighet i et forsøk uten tilbakelegging
14 Beregne betinget sannsynlighet i en asynkron situasjon

Andre oppgaver som måler samme enheter, som oppgave 2 og oppgave 7, viser i motsetning til oppgave 6a, at undervisning nytter.

La oss se nærmere på for eksempel oppgave 1a. Der viser endringen en økning på noe over 8 prosent. Det mener jeg er en mye sterkere positiv tendens enn for eksempel oppgave 3 med en økning på vel 11 prosent. I oppgave 1a har om lag halvparten av elevene som ikke fikk til denne oppgaven før undervisning, fått til oppgaven etter undervisning. Oppgave 3 kan ikke vise til en tilsvarende forbedring, og selve prestasjonen på oppgaven er fremdeles lav – godt under 50 prosent. På samme måte vil jeg si at oppgave 7 har en god positiv tendens, selv om dette ikke er vist med grønn farge på oppgaven. Tall fra CPR bekrefter dette, med en økning på 14 prosent. Vi må altså ikke se oss blinde på andelen som har klart oppgavene, vi må også se på dem som ikke har klart oppgavene. Jeg har ikke markert oppgave 7 som en grønn oppgave på grunn av det negative resultatet i kolonne 7 (-4,4 %). Det er mulig å kalibrere alle prestasjoner i hver enkelt oppgave til en middelverdi, for eksempel 50 prosent. Jeg synes selv at det er å gå litt langt og dersom en oppgave er veldig skjev (høy eller lav score) vil hver enkelt liten endring kunne gi store prosentvise utslag. Jeg velger blant annet derfor og av behagelighetshensyn både for meg og leser, å la tallene stå slik de fremkommer i tabellene ovenfor.

Etter å ha sett på alle resultatene og alle oppgavene – hvem var det som var best av alle de ti gruppene? Jeg beregner en gjennomsnittscore for hver gruppe. Siden Diaz med flere har beregnet samlet poengsum på oppgave 8, multiplisere jeg Diaz med flere sin score på denne oppgaven med 2. Und4 hadde ikke alle oppgavene og beregning av snitt kan da falle uheldig ut, både den ene og den andre veien. Resultatet er derfor justert ved direkte sammenligning, oppgave for oppgave, med Und1. Den gjennomsnittlige prestasjonen for hver gruppe:

pilot - F	und 1 - E	und2 - E	und3 - E	und4 - F	und5 - F	Und6 - E	D&B - F	D&B - E	D&F - E
58,7	67,5	48,4	54,0	51,0	48,8	59,1	37,8	57,1	52,0

Tabell. Justert gjennomsnitt i prosent for alle gruppene.

Vi ser at Sam1 og Sam2 har en forbedring i resultat. Vi ser også at S1-gruppene ikke ligger så langt bak de andre som har hatt undervisning, men gjennomsnittlig ligger de lavere enn vg1-gruppene uten undervisning. Jeg mener at jeg med dette kan forsvare min bruk av Sam1 og Sam2 (og utelatelsen av S1-gruppene) i analysen. Vi kan stille spørsmålet om S1-gruppene burde vært med i det hele tatt, men da sier jeg ja – blant annet med henvisning til to omvendt betingede sannsynligheter i oppgave 5 og 6b.

4.16 Bekreftelse på validitet

Sentralmål og spredningsmål i CPR-testen bekreftet testens validitet, altså evnen til å skille studenter mellom de to gruppene «Før» og «Etter». Krav til validitet:

- Sentralmålene er forskjøvet mellom «Før» og «Etter»-gruppene.
- Spredningsmålene, standardavvik og boks-plot, er noenlunde identiske for gruppene.

CPR-testen har validitet og Diaz og Batanero kan konkludere med at undervisningsøkten har hjulpet.

Tabellen fra kapittel 4.14, nå også med mine resultater, trenger noen presiseringer: Poengsummen i mine undersøkelser var 18, mens poengsummen i CPR var 30. Alle mine tall i tabellen nedenfor er justert lineært fra 18 til 30 for å kunne foreta en sammenligning mot CPR. Denne justeringen er i seg selv ikke nødvendig og ikke helt korrekt prosedyre, men det letter lesbarheten. Jeg har kun tatt med de fire gruppene fra direkte sammenligning (Sam1 og Sam2).

Gruppe	antall	Gjennomsnitt	standardavvik	Q1	Median	Q3
D & B - F	177	12,68	5,69			
D & B - E	206	18,47	5,15			
Pilot og Und5 (F)	44	15,72	4,54	11,67	16,67	18,33
Und1 og Und6 (E)	36	18,70	4,12	16,67	19,17	21,67

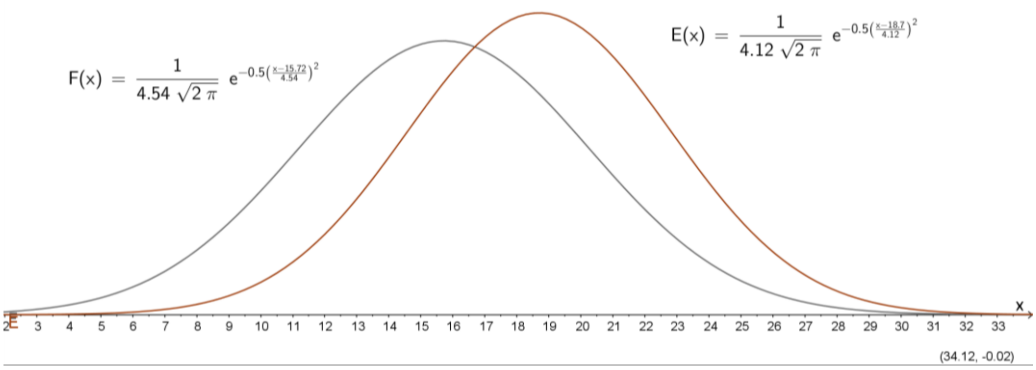
Tabell. Grupper sortert i Før og Etter. Antall personer. Sentral- og spredningsmål.

Finner jeg tilsvarende resultat og kan jeg konkludere på samme måte? Tabellen ovenfor viser:

- Nedre og øvre kvartil og median ligger alle høyere for Und1 og Und6 (Etter)
- Sentralmålene er altså forskjøvet mellom gruppene til fordel for Und1 og Und6
- Spredningsmålene, standardavvik og boks-plot, er noenlunde identiske for gruppene.

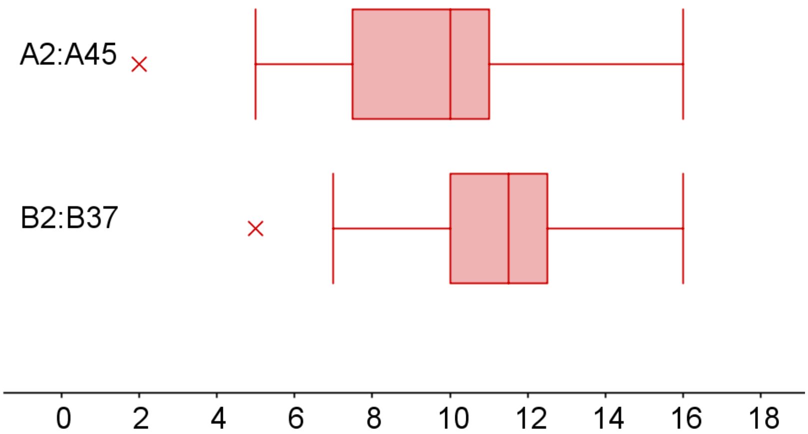
Ut fra dette kan jeg gjenta, siden det allerede er sagt i kapittel 3, at den kvantitative delen av min undersøkelse har god validitet. Jeg har riktignok færre deltakere, men til gjengjeld er standardavviket mindre.

Ved å anta en normalfordeling av datamaterialet i mine undersøkelser kan vi illustrere forskjellen. Jeg har brukt skalert gjennomsnitt og standardavvik fra tabellen ovenfor.



Figur. Grafer til normalfordelingskurver med skalert gjennomsnitt og standardavvik. Direkte sammenligning av fire grupper fordelt i «Før» og «Etter».

Batanero og Diaz kunne fastslå med en sikkerhet på 82,3 prosent om tilfeldig student tilhørte den ene eller den andre gruppen. Jeg har ikke beregnet denne sikkerheten for mine undersøkelser, men konstaterer at avstanden mellom mine to gjennomsnittsverdier (15,72 og 18,70) er mindre enn standardavvikene. For Diaz og Batanero er dette omvendt. Det betyr at jeg med mindre grad av sikkerhet kan si om en tilfeldig valgt deltaker tilhører den ene eller den andre gruppen. Men kurvene ovenfor og boksplot nedenfor tilsier at vi kan skille mellom gruppene. Forskyvningen av kvartiler og median er tydelige, men formene er ikke identiske. Forklaringen på formen ligger i at tallmaterialet er lite og ved å studere tallene nøye, observerer jeg at vi er én plassering fra å forskyve Q1 til høyre på den øverste boksplotten og til venstre på den nederste boksplotten. De er altså ikke så ulike i form som det kan virke som. Boksplot forholder seg til den totale poengsummen på 18, og er ikke skalert opp til en poengsum på 30 som i tabellen ovenfor. Dette kan vi neglisjere ved å overse tallinja nederst. Kryssene til venstre er uteliggere. Uteliggere er minimums- eller maksimumsverdier som er så ekstreme at de ikke regnes med i datamaterialet.



Figur: Boksplot.»Før» øverst og «Etter» nederst.

Vi kan lese at høyeste score for begge gruppene (Før og Etter) var 16 poeng av 18. Dette forklarer også de ulike lengdene (whiskers) fra øvre kvartil til maksimum på høyre side.

5 Drøfting

Resultatene fra CPR-testen til Batanero og Diaz viser generelt god forbedring i problemer med betinget sannsynlighet og formell forståelse av betinget sannsynlighet ved hjelp av undervisning. De observerte høyere score i instruksjonsgruppen på alle områder som omhandler formell matematisk kunnskap. Mine resultater i kapittel 4 samsvarte i stor grad med Batanero og Diaz sine resultater og jeg konkluderte med at undervisning hjelper, og svarte da på første del av forskningsspørsmålet. De neste spørsmålene dreier seg om misoppfatninger og feilslutninger:

- **Vil undervisning redusere antall misoppfatninger?**
- **Hvilke misoppfatninger består, øker eller minker etter undervisning?**

I dette kapitlet drøftes resultatene i lys av misoppfatningene beskrevet i kapittel 2. Tabellen nedenfor viser misoppfatningene koblet mot oppgaver slik det er gjort i dette kapitlet. Noen oppgaver avslører bestemte misoppfatninger, mens andre oppgaver kan lede til flere ulike misoppfatninger. Jeg har noen ganger trukket inn andre oppgaver som støtte for en misoppfatning. De oppgavene er i kolonnen til høyre. For eksempel blir misoppfatning 1 målt på bakgrunn av oppgave 1, men jeg søker støtte for misoppfatningen i oppgave 11 og 12.

Nr	Misoppfatning fra kapittel 2	Oppgave	Støtte
1	Kontekst og forveksling	1	11 og 12
2	Snittet mellom to hendinger, conjunction fallacy	3	
3A	Motsatte betingede sannsynligheter - forveksling og likhet	4	1cd
3B	Motsatte betingede sannsynligheter - inversjon av tidsaksen	5 og 6b	
3C	Motsatte betingede sannsynligheter - bayesiske problemer	13	
4	Synkrone og asynkrone hendinger	2, 6a, 7, 11, 12	
5A	Heuristikker - Siste observasjoner og representativitet	10	
5B	Heuristikker - Base Rate		11 og 12
6A	Den mest vanlige - bestemt sannsynlighet	9 og 10	
6B	Den mest vanlige - lik sannsynlighet	9 og 10	2
6C	Den mest vanlige - bestemt regel	9	
7	Kombinatoriske problem	8	9
8A	Heuristikker - tilgjengelige utfall		9 og 10
8B	Heuristikker – utfallstilnærming		10
Syst1	System 1-feil. Beskrevet i kapittel 1	9	3

Tabell: Misoppfatninger fra kapittel 2: Nummer, beskrivelse og oppgaver for misoppfatning.

Når samlet resultatet fra kapittel 4 viser fremgang i prestasjonene med undervisning, må vi også konkludere med at undervisning gir en samlet reduksjon i antall feilslutninger og misoppfatninger. Men dette resultatet forteller ikke alltid så mye om den enkelte misoppfatning fra listen ovenfor. Elevene gjør mange ulike feil, og det er nødvendigvis ikke slik at feilene kommer som en konsekvens av en spesifikk misoppfatning. Drøftingen videre i dette avsnittet vil dreie seg om en sikrere spesifisering rundt hver enkelt misoppfatning fra tabellen ovenfor.

Det er vanskelig å ha tette skott mellom misoppfatninger, enheter og oppgaver, og dette kapitlet har derfor en inndeling som er inspirert av metodekapittelet; blandet metode. Kapitlet er først organisert etter misoppfatningene fra kapittel 2 og deretter organisert etter oppgavene fra kapittel 4. Det forekommer at et enkelt svar i en oppgave kan ha sitt opphav i mange ulike misoppfatninger, men hvilken spesifikk misoppfatning hadde eleven egentlig? Jeg skal være forsiktig med å trekke for mange veksler på et enkelt svar. Balansen mellom å åpne opp for misoppfatninger og samtidig snevre inn blir viktig. Som i kapittel 4 vil jeg trekke inn og ut de resultat og svar som jeg finner formålstjenlig for mine drøftinger. Oversikten viser hvilke misoppfatninger som blir omtalt hvor i dette kapitlet. Derfor er også oversikten over avsnittene i kapitlet fjernet.

Kapittel 5	**Omtalt misoppfatning**
5.1	1
5.2	2
5.3	3A, 3B, 3C
5.4	4, 6B, 1, 8C, 6C
5.5	7
5.6	7, 6A, 6C
5.7	5, 6A, 8A, 8C
5.8	6B

Tabell. Kapittelinndeling og misoppfatning fra kapittel 2.

5.1 Kontekst

Avsnitt 2.1 dreier seg om misoppfatninger relatert til kontekst. Oppgave 1 i testen målte i hvilken grad undervisning hjelper på å tyde og forstå teksten. Det er eksempler på at elever forveksler sannsynlighetene $P(A|B)$, $P(\mathrm{B}|\mathrm{A})$ og $P(A \cap B)$. Tabellen viser hvordan elever før og etter undervisning forveksler de ulike sannsynlighetene.

Oppgave	a	b	c	d			
Hending	H	H og E	H|E	E|H	E		
Svar	104/780	75/780	75/350	75/104	350/780	104/350	annet
Før	**65**						
a	56	0	1	0	0	0	8
b	1	41	9	2	1	0	11
c	1	4	46	1	1	0	12
d	0	3	2	42	2	6	10
							41
Etter	**71**						
a	67	0	0	0	0	0	4
b	1	46	8	8	1	1	6
c	0	3	59	0	1	0	8
d	0	1	4	53	3	6	4

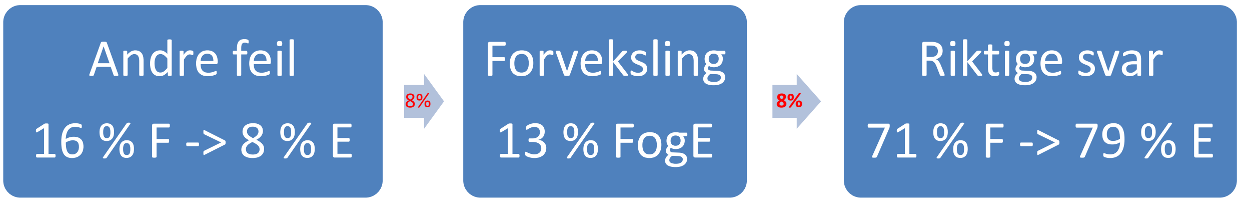
Tabell. Oppgave 1: Forveksling mellom sannsynligheter før og etter undervisning.

Eksempler til hvordan tabellen skal leses:
Det er 65 «før»-elever og 71 «etter»-elever. Antall elever som har svart riktig på hver enkelt oppgave finner vi i de gule cellene som går diagonalt. Oppgavene (a, b, c og d) finner vi i første kolonne (grå), og svarene som er gitt til hver oppgave starter i kolonne 2. For eksempel viser den nederste rekken at på oppgave 1d har 4 av 71 «etter»-elever svart med den omvendte sannsynligheten (brukt svaret fra oppgave 1c) og 53 har svart riktig. De mest fremtredende resultatene er at:

1. Mange elever før (9 av 65) *og* etter (8 av 71) undervisning forveksler $P(H \cap E)$ med $P(H|E)$, rød farge.
2. Flere elever (8 av 71) *etter* undervisning forveksler $P(H \cap E)$ med $P(E|H)$, grønn farge.
3. Andel andre feil er kraftig redusert etter undervisning (siste kolonne).
4. En elev hadde svaret fra 1d, $P(E|H)$, på de tre oppgavene 1b, 1c og 1d.
5. Forveksling mellom de to betingede sannsynlighetene holder seg konstant, blå farge. Dette omtales videre i avsnitt 5.3.1.

Til oppgave 1d finner vi en annen interessant observasjon. Vi kan se at det er mange som svarer at sannsynligheten er 104/350; de gunstige er alle som tidligere har hatt hjertestans, og de mulige er alle som er over 55 år. At det er elementer som er gunstige, men ikke mulige, ser ikke ut til å bekymre elevene nevneverdig. Jeg kaller dette for en «kvasibetingelse», og vi skal møte det igjen. Det er kun intervjugruppe 3 og enkeltintervju 4 som diskuterte denne oppgaven. I intervjugruppe 3 hersket det en del forvirring rundt oppgave 1b og 1d. På direkte spørsmål fra meg om oppgave 1b og 1d er samme spørsmål, svarer de med en gang at spørsmålene er ulike – ingen tvil om det. Etter en kort samtale elevene imellom finner de begrensningen i utfallsrommet.

Resultatene i punkt 1 og punkt 2 ovenfor samsvarer med Falks (1989) mening om at ordlyden av betinget sannsynlighet, eller konteksten, i stor grad gjør det vanskelig for elevene å skille mellom betingede hendinger og snittet mellom hendinger (Diaz og Fuente, 2007). Generelt viser resultatet fra kapittel 4 viser en forbedring på 8 % etter undervisning. Et dypdykk i dataene viser likevel at andelen forvekslinger mellom de ulike sannsynlighetene som beskrevet i kapittel 2.1, holder seg konstant med 13,1 % (før) og 13,0 % (etter). Forbedringen skyldes altså at elevene i mindre grad gjør «andre» feil. Jeg har liten mulighet til å kartlegge hvilke elever som forflytter seg hvor, men jeg mistenker at de fleste (8 %) rykker opp en divisjon slik figuren nedenfor viser.



Figur. Oppgave 1cd: Andelen feil, deriblant forveksling, og riktige svar «før = F» og «etter = E». Andelen «andre feil» reduseres med 8 % og «riktige svar» øker med 8 %. «Forveksling» er konstant.

Andre oppgaver som kan bekrefte Falks påstander er oppgave 11 og 12. I begge oppgavene skal elevene beregne sannsynligheten for snittet $P(A \cap B)$, men flere elever beregner sannsynligheten til en kvasibetingelse av $P(A|B)$. De beregner forholdet, $P(A)/P(B)$, heldigvis med den største sannsynligheten i nevneren. Og heldigvis igjen; andelen beregninger av denne typen reduseres med undervisning. Konklusjonen er at oppgave 1 viser at elevene har klart å knekke «tekst»-koden og at det skjer mindre grad av feil etter undervisning.

5.2 Snittet mellom to hendinger

Misoppfatningen om at snittet mellom to hendinger er større enn en av hendingene, conjunction fallacy (kapittel 2.2), måles i oppgave 3. Andelen elever som har denne misoppfatningen er motsatt av andelen som svarer riktig. Vi kan altså bruke resultatet fra kapittel 4 indirekte. Vi har en endring i prestasjonene fra 25 prosent til 29 prosent før og etter undervisning for alle grupper. Selv om tallene viser at undervisning hjelper, er forbedringen marginal på et allerede lavt utgangspunkt. Intervjuene viste at det var stor avstand mellom oppgaven og forståelsen til elevene.

I intervju 1 volder oppgaven store problemer. De er rimelig overbevist om at alternativ c er det riktige svaret; begge alternativene a og b er like sannsynlige. De begrunner det med at den første opplysningen ikke virker inn på om spilleren vinner kampen (må vinne tre sett), fordi det gjenstår fire sett og det er 50 % sjanse for å vinne to av disse. De konkluderer med at selv om du vinner det første settet, kan du tape eller vinne kampen, fordi det er de to mulige utfallene; altså 50/50, og overfører dette direkte til alternativ c, som om spørsmål a inneholdt en betingelse om at spilleren vant det første settet **og** kampen. Elevene i denne gruppen roter mye frem og tilbake og kommer ikke frem til det riktige svaret uten sterk veiledning fra meg. De holdt på alternativ c så lenge de kunne. Selv den ene eleven, som hadde svart riktig under testen, gikk vekk fra dette og holdt sterkt på alternativ c. Jeg vil konkludere med at det var vanskelig å rokke ved deres tro, selv etter at jeg hadde guidet dem mot målet. Jeg følte at de ikke var helt overbevist.

I intervju 3 sier en elev at alternativ c er det riktige svaret. Jeg spør om hun da mener at alternativ a og b er like sannsynlige? Da tenker hun seg litt om før hun svarer:

«Hm-hm. Nei vent. Det var sånne rare alternativer her.»

En annen elev i gruppen mener at a er mest sannsynlig på grunn av at:

«På a er det 50/50, mens på b der er det en vei han må gå».

Hun sier at hun mener at valgtreet er litt lengre på b – jo, lengre tre, jo mindre sannsynlig. Hun er også den eneste av de tre elevene som har svart riktig på oppgaven, men det finnes også andre lyspunkt: eleven i intervju 4 er urokkelig i sin tro på det riktige alternativet a.

Forskningen viser også til store problemer med denne misoppfatningen. En undersøkelse utført av Fishbein og Schnarch (1997) viste en forbedring i prestasjon i conjunction fallacy med økende alder hos elever, men et annet arbeid utført av Davidson (1995) viste det motsatte (side 62 i Watson og Moritz, 2002). Økende alder og undervisning er ikke sammenfallende med økende prestasjoner i problemer som involverer conjuction fallacy. Selv om det generelt blir en bedre prestasjon dersom opplysningene i problemene er gitt på frekvens-form, var dette ikke tilfelle ved conjunction fallacy (side 80 i Watson og Moritz, 2002).

Min konklusjon er at resultatene og intervjuene bekrefter at «conjunction fallacy» står sterkt, både før og etter undervisning. Undervisning fører til en liten forbedring.

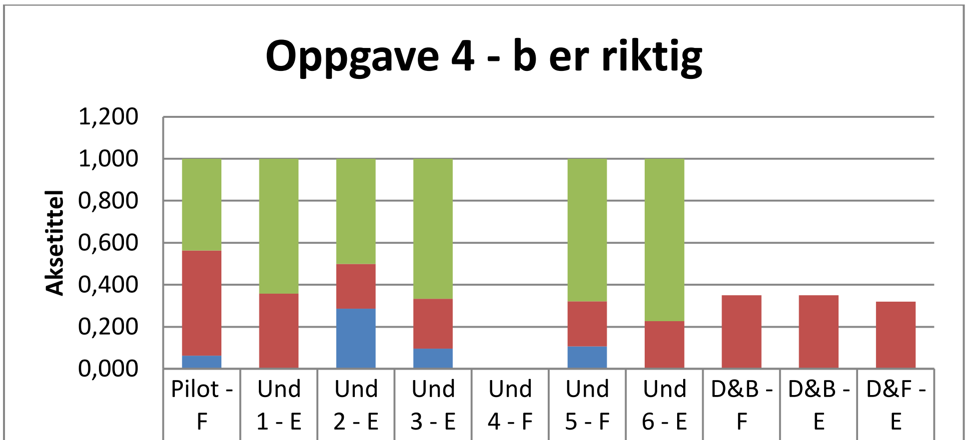
5.3 Motsatte betingede sannsynligheter

I kapittel 2.3 ble teorien rundt misoppfatninger mellom to motsatte betingede hendinger beskrevet. Hvilke av disse misoppfatninger finner jeg i mine undersøkelser og hvilke oppgaver kan gi svar av betydning?

5.3.1 Forveksling og likhet

Oppgave 4 måler i hvilken grad elevene er i stand til å identifisere to motsatte betingede sannsynligheter. Resultatet fra mine undersøkelser viste at denne misoppfatningen stod sterkt, og faktisk hadde en negativ trend etter undervisning, fra 32 til 25 prosent for alle mine grupper.

Diagrammet nedenfor viser fordelingen på de ulike alternativene. På toppen er alternativ c, mens det riktige alternativet b (Positivt utslag|Kreft) er i midten og markert med rød farge. På multiple choice-oppgavene har jeg kun tilgang til de riktige svaralternativene fra Diaz med flere.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E	D&B - F	D&B - E	D&F - E
c) Like	0,438	0,643	0,500	0,667	0,000	0,679	0,773	0,000	0,000	0,000
b) P|K	0,500	0,357	0,214	0,238	0,000	0,214	0,227	0,350	0,350	0,320
a) K|P	0,063	0,000	0,286	0,095	0,000	0,107	0,000	0,000	0,000	0,000

Figur: Oppgave 4. Prosentandel for alle alternativer: a, b og c. F = Før og E = Etter.

Ved å studere diagrammet, og spesielt alternativ c, ser vi at elevene i stor grad oppfatter to motsatte betingede sannsynligheter som to sider av samme sak, uavhengig av undervisning. Und2 skiller seg litt ut ved at denne gruppen er den eneste som gir større sannsynlighet til alternativ a enn alternativ b. Det var over 60 prosent av alle elevene i mine undersøkelser som svarte alternativ c, at de to hendingene er de samme. Dette bekrefter at de har store problemer med å se forskjell på to omvendt betingede sannsynligheter.

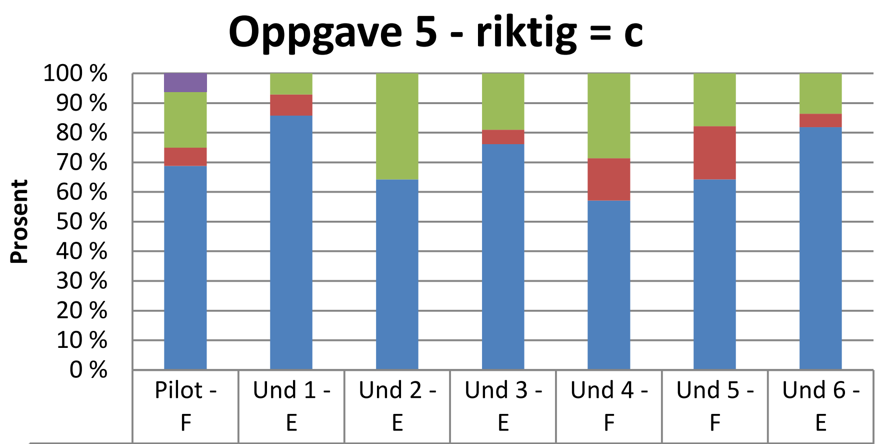
Denne oppgaven er kun diskutert i det semistrukturerte intervjuet med én elev. Eleven bekreftet teorien om at elever har problemer med å skille to omvendt betingede sannsynligheter.

> *«Jeg ser ikke helt forskjellen på a og b, bortsett fra at de står i annerledes rekkefølge. At det er positivt når personen har kreft og at personen har kreft når testen er positiv. Jeg ser ikke helt forskjellen. Så jeg mener c.»*

Misoppfatningen mellom to betingede hendinger (The fallacy of the transposed conditional) står sterkt i oppgave 4. Forskning viser også at elever og studenter forveksler to omvendt betingede sannsynligheter (side 295 i Falk, 1986). Resultatet samsvarer også med Diaz og Fuente, der 59 prosent av studentene i den første undersøkelsen (med undervisning) ikke så forskjell på en betinget sannsynlighet og dens inverse. Når elevene skal finne betingede sannsynligheter fra en krysstabell, oppgave 1c og 1d, viste mine undersøkelser at forvekslingen mellom to motsatte betingede sannsynligheter holdt seg konstant (se tall fra tabell i kapittel 5.1). Konklusjonen er at elever har vanskeligheter med å skille to omvendt betingede sannsynligheter fra hverandre, både før og etter undervisning.

5.3.2 Inversjon av tidsaksen

Det er to oppgaver som viser seg til misoppfatningen om at betingelse må komme før effekt; oppgave 5 og 6b. Vi ser først på oppgave 5. I forrige kapittel observerte vi en reduksjon i prestasjon etter undervisning i mine grupper, mens Diaz og Batanero hadde en svak økning. Den gjennomsnittlige scoren for alle grupper er likevel lav, om lag 17 prosent. At mange har problemer med tilbakedatering av betingelser bekreftes ved at gjennomsnittlig 71 prosent av elevene velger alternativ a (1/2 – uten betingelse). Tabellen nedenfor viser den prosentvise fordelingen på de ulike alternativene.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E
d) um	0,063	0,000	0,000	0,000	0,000	0,000	0,000
c) 2/3	0,188	0,071	0,357	0,190	0,286	0,179	0,136
b) 1/3	0,063	0,071	0,000	0,048	0,143	0,179	0,045
a) 1/2	0,688	0,857	0,643	0,762	0,571	0,643	0,818

Tabell: Andel alle alternativer oppgave 5. Alternativ c er riktig.

I intervjusituasjonen volder oppgave 5 store problemer, men alle elevene mener at oppgaven er løsbar. Alle som er i gruppeintervju mener at det riktige svaret er ½, og en person i det første intervjuet nekter helt å forholde seg til en betingelse som kommer etter effekten. Han har veldig lyst til å snu situasjonen og når vi nærmer oss løsningen på oppgaven, sier han:

> *«Det er et dumt spørsmål. Det skulle ikke vært lov å stille sånne spørsmål - det er min mening. Gitt at ballen går gjennom R, hva er sjansen for at den går gjennom 1 eller 2? Det snur opp ned på hele spørsmålet – totalt. Det er et misledende spørsmål og de vil svare feil.»*

I intervjugruppe 2 spør jeg hva som er vanskelig med oppgaven. En elev svarer slik:

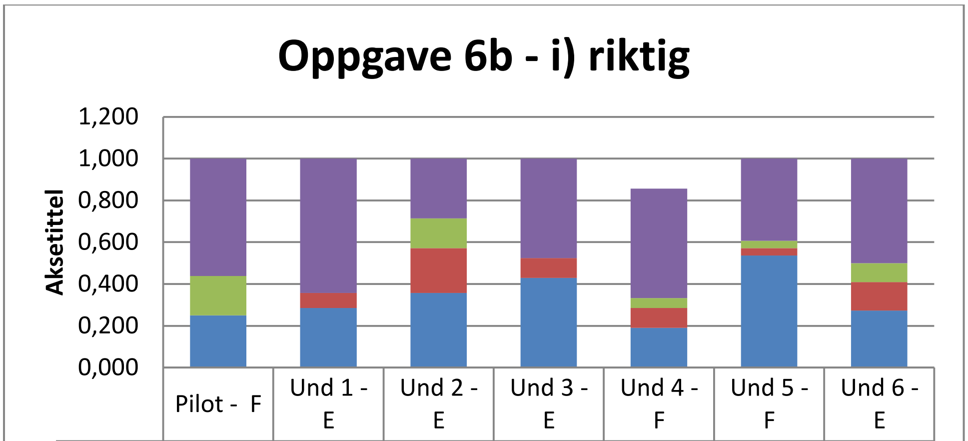
> *«Jeg tenkte at det var veldig vanskelig å finne ut av hvor den landet etterpå, når du ikke visste hvor den landet først.»*

Denne eleven beskriver altså problemet med rekkefølgen av to hendinger. I den siste intervjugruppen er det en elev som evner å se situasjonen ved å si at det er tre mulige utfall med den gitte betingelsen. Og med denne oppskriften overbeviser hun de to andre. Selv om elevene svarte feil på testen, resonnerte de seg frem til det riktige svaret. Som en dem sa: *«Man må tenke»* - eller koble inn system 2? I det semistrukturerte intervjuet er det ingen tvil. Enkel resonnering og eliminering av feilaktige svar leder til det riktige svaret 2/3.

I oppgave 6b viser både mine og Diaz og Batanero sine resultater en reduksjon i prestasjon etter undervisning, men den gjennomsnittlige scoren på 32 prosent for alle grupper, er vesentlig høyere enn for oppgave 5. Gir oppgave 6a en pekepinn, eller et varsko, om at her må du tenke deg nøye om – koble inn system 2? Ulike misoppfatninger knyttet til ulike svaralternativer

i) 1/3 - Det riktige svaret
ii) Umulig å avgjøre. Nekter å forholde seg til tilbakedatering av informasjon
iii) 1/6 - Bruker produktsetningen der denne ikke skal brukes.
iv) ½ - Neglisjerer betingelsen på grunn av manglende kompatibilitet med tidsaksen.

Diagrammet nedenfor viser at dersom en elev ikke velger det riktige alternativet i, går eleven for at en betingelse ikke har tilbakevirkende kraft, og dermed alternativ iv. De sidestiller $P(W_1|W_2)$ med $P(W_1)$. Det siste resultatet er i samsvar med undersøkelsen til Falk på 88 universitetsstudenter. (Falk, 1989, side 130 i Diaz og Fuente, 2007). Vi ser også at Und4 hadde noen ubesvarte oppgaver.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E
iv) 1/2	0,563	0,643	0,286	0,476	0,524	0,393	0,500
iii) 1/6	0,188	0,000	0,143	0,000	0,048	0,036	0,091
ii) um	0,000	0,071	0,214	0,095	0,095	0,036	0,136
i) 1/3	0,250	0,286	0,357	0,429	0,190	0,536	0,273

Tabell: Andel alle alternativer oppgave 6b. Und4 har noen ubesvarte oppgaver.

I intervjusituasjonene brukte vi lang tid på oppgave 6b, og alle gruppene diskuterte denne oppgaven. I intervju 1 brukte vi 8 minutter på oppgave 6b. Det første en gutt sier (nærmest et utbrudd) er:

> *«Dette spørsmålet er utrolig. Gitt at det som skjer etter det vi spør om… Jeg vil si at den som kommer etterpå skal ikke ha noe å si for du har fortsatt femti prosent sjanse når du har alle fire. Sånn at det er en halv.»*

Gutten som sier dette er den samme som protesterte på oppgave 5. En annen elev følger opp:

> *«Uansett om du trekker en hvit etterpå så har du to hvite og to svarte i begynnelsen.»*

Den tredje eleven sier:

> *«Jeg tenker det er en halv, men jeg tror ikke det er riktig.»*

Hun baserer dette på at spørsmålet måtte vært stilt annerledes dersom ½ var det riktige svaret. Etter en lang diskusjon nærmer vi oss det riktige svaret og da utbryter den første gutten:

> *«Disse spørsmålene er dumme. Ikke overraskende at vi tar feil hver gang. Dette er bare dumt.»*

Mens den tredje eleven sier, fornuftig nok (skulle tro at hun hadde lest Falk):

> *«Poenget er at dersom du får vite noe, så…, så må du bruke det.»*

I intervju 2 brukte vi 10 minutter på denne oppgaven. Den var vanskelig å drøfte, både for elevene og meg. En elev sier at denne oppgaven «var litt verre». Hun peker på at tidsaksen og rekkefølgen skaper en uorden som ikke var til stede i oppgave 6a. Elevene ber om muligheten til å tegne et valgtre og starter en diskusjon basert på diagrammet. Hun foreslår at de må «regne baklengs». De ender opp med 1/3 og 2/3, men så nær målet stopper det plutselig opp, og elevene beveger seg i samlet flokk tilbake til alternativ iv: 1/2. Jeg geleider dem tilbake til nest siste stopp ved å påpeke ulikheten mellom dette spørsmålet og spørsmålet som ville gitt svaret 1/2. Først etter en god diskusjon klarer de å godta at en hvit kule er skjermet til andre trekning, på samme måte som en hvit kule var skjermet til første trekning. Det er ingen forskjell og vi kunne faktisk ha trukket begge kulene samtidig. En elev avslutter med å si, litt oppgitt og til latter fra de andre:

> *«That makes so much sense.»*

I intervju 3 brukte vi 7 minutter på denne oppgaven. Det første som blir sagt er:

> *«Den er irriterende, fordi man kan på en måte ikke tenke slik som man pleier å gjøre.»*

De finner noenlunde fort ut at situasjonen er den samme som i 6a.

> *«For du gjør jo akkurat det samme som i sted. Du trekker to kuler, og så ser du på en av dem. Tingen er at du bare ser på den siste du trakk.»*

Etter litt diskusjon frem og tilbake, konkluderer de med at den gitte informasjonen innvirker på sannsynligheten. Den ene hvite kulen er reservert til siste trekning, men det riktige svaret 1/3 sitter likevel langt inne.

I intervju 4 brukte vi ikke så lang tid. Det kan være fordi den enslige eleven ikke har så mange andre å diskutere med. Han er urokkelig i sin tro på at hendinger kommer i en bestemt rekkefølge. Da jeg påpeker at det er en hvit igjen til andre trekning svarer han:

> «: Ja, men det kan du ikke ta hensyn til. Her står det at «Hvor stor er sannsynligheten for at den **første** kulen er hvit, ikke den andre».»

Eleven lar seg overhodet ikke overbevise. Det synes nytteløst å gå imot tidsaksen.

Å forstå at sannsynligheten til en hending kan endres i takt med nye opplysninger som stilles til rådighet senere, kan synes vanskelig. Dette burde ikke være et problem, mener Falk og viser til medisin. Vi vet alle at sykdommen (årsaken) kommer først, og symptomene (effekten) kommer etterpå, men vi kan likevel hekte sannsynligheter til en sykdom med kunnskap i symptomene (side 294 i Falk, 1986).

Hvordan ville oppgaven vært uten de fire svaralternativene? Det er et interessant spørsmål. Jeg tror at vi i stor grad ville fått de samme svarene som tidligere, med noe større oppslutning rundt ½ og 1/6, mens alternativet 1/3 ville fått lavere oppslutning. Ved å bruke eliminasjonsmetoden tenker jeg at noen elever har svart 1/3 uten å være hellig overbevist. Intervjusituasjonene viser også dette. Det kan jo til dels skyldes min veiledning i intervjuene, men samtidig vil noen elever ledes mot 1/3 ved hjelp av eliminasjon. Misoppfatningen med betingende sannsynligheter og motsatt tidsakse består i beste velgående og øker med undervisning. Det kan virke som om det positive arbeidet med betinget sannsynlighet, valgtrær og produktsetningen (se kapittel 5.4) virker negativt på evnen til å godta inversjon av tidsaksen.

5.3.3 Bayesiske problemer

Oppgave 13 måler graden av bayesiske problemer beskrevet som misoppfatning 3C i kapittel 2. I mine undersøkelser fikk jeg liten til ingen informasjon, og gjør som jeg sa i kapittel 4. Jeg holder meg unna drøfting av denne oppgaven, men jeg skulle likt å ha sett den presentert med frekvenser.

5.4 Synkrone og asynkrone hendinger

I hvilken grad elevene behersker betinget sannsynlighet og produktsetningen i synkrone og asynkrone hendinger, finner vi svar på i mange oppgaver, blant annet oppgave 2, 6a, 7, 11 og 12. Oppgavene er med og uten tilbakelegging. Hver oppgave er kommentert med bakgrunn i resultat i kapittel 4, her vil vi i se på dem samlet. Oppgave 2 og 6a er direkte spørsmål om betinget sannsynlighet, mens oppgave 7 er en tilnærming til produktsetningen som brukes i oppgave 11 og 12.

Oppgave	2	6a	7	11	12
Pilot - F	0,813	0,875	0,813	0,500	0,500
Und1 - E	0,857	0,786	0,786	0,714	0,714
Und2 - E	0,714	0,786	0,500	0,571	0,429
Und3 - E	0,857	0,810	0,762	0,714	0,714
Und4 - F	0,000	0,810	0,000	0,667	0,619
Und5 - F	0,857	0,786	0,821	0,321	0,321
Und6 - E	0,909	0,727	0,955	0,591	0,591
D & B - F	0,770	0,720	0,620	0,254	0,237
D & B - E	0,890	0,810	0,760	0,490	0,617
D & F - E	0,890	0,690	0,760	0,520	0,600

Tabell. Samlet resultat oppgave 2, 6a, 7, 11 og 12 for alle grupper. Andel riktig svar.

Alle resultater over eller lik 60 prosent er markert med grønn farge. Vi ser at på multiple choice-oppgavene (2, 6a og 7) scorer elevene over dette, med ett unntak (rød farge). Gjennomsnittscore for multiple choice – alle grupper:

- Oppgave 2: 84 %
- Oppgave 6a: 78 %
- Oppgave 7: 75 %

På oppgave 11 og 12 ser vi en markant forbedring etter undervisning. Dette er oppgaver som behandler synkrone og asynkrone hendinger med og uten betingelser. Produktsetningen er sentral og vi ser at elevene i stor grad behersker dette. Forbedringen er på hele 18 prosent, fra 43 til 61 prosent, men vi ser også på Und5 at elevene nødvendigvis ikke behersker produktsetningen intuitivt (rød).

Hvilke misoppfatninger kan vi observere? I oppgave 2 er det bare én av alle elevene som har svart alternativ i. Det er grunn til å tro at elevene som har svart alternativ iii ikke har oppfattet at situasjonen er uten tilbakelegging. Snarere det alternativet enn at det er to utfall igjen, og sjansen deles likt mellom disse, omtalt som 6B Lik sannsynlighet i kapittel 2. Men sikker kan jeg ikke være. I oppgave 6a er gjennomsnittscoren for alle grupper på 78 prosent, altså et meget godt resultat. Dette medfører at misoppfatningene som forekommer i mine resultater er få, men de som dukker opp er forvekslingen mellom spørsmålet $P(W_2|W_1)$ og svaret $P(W_1 \cap W_2)$, både med og uten tilbakelegg. Slike forvekslinger er omtalt i avsnitt 5.1 Kontekst.

I oppgave 7 gjelder det å kunne utarbeide et fullstendig utfallsrom og å kunne bruke produktsetningen. De elevene som svarer feil, har en større tendens til å velge alternativ b fremfor alternativ a. Mest sannsynlig bommer de på dette med tilbakelegging eller ei, altså en forveksling av sannsynligheter. Elevene uten undervisning viste gode prestasjoner i disse oppgavene, noe som viser at de har en intuitiv forståelse av sannsynlighet med synkrone og asynkrone hendinger. Det gode resultatet før undervisning blir et hinder for å observere en eventuell økning. Mine konklusjoner avviker derfor noe fra Batanero og Diaz sine. I mine undersøkelser kan jeg konkludere med at prestasjonen er god for både «før»- og «etter»-gruppene. Batanero og Diaz har lavere score på «før»-gruppen og kan derfor vise til en større forbedring for «etter»-gruppen. Dette understreker at undervisning hjelper.

I oppgavene 11 og 12 skal vi bruke produktsetningen. Begge oppgavene fikk en markant forbedring etter undervisning. Hvilke feil gjøres? Hvilke misoppfatninger har elevene – før og etter undervisning? Oppgavene er såkalte åpne oppgaver som åpner opp for delvis riktige svar. I oppgave 11 argumenterer Batanero og Diaz for at studenter som gjenkjenner tallene og bruker disse skal kategoriseres som delvis riktig svar. De viser til eksemplet: 0.70 + 0,80 = 1,50. Da tenker jeg at det er mange mulige alternativer til delvis riktig svar. Jeg har brukt Batanero og Diaz som utgangspunkt for å kategorisere et svar som delvis riktig eller feil, men samtidig har jeg vært nødt til å kategorisere svar som jeg ikke fant omtalt.

Kommentar	Nr	svar	Kategori	Før	Etter
Forveksling på grunn av kontekst. Ikke produktsetningen.	1	70/80	Delvis	5	3
I nærheten av 0,56	2	ca 50 %	Delvis	7	3
Tar den minste sannsynligheten. Overser base rate?	3	70 %	Feil	4	7
Gjenkjenner data. Tar gjennomsnittet.	4	75 %	Feil	7	1

Tabell. Oppgave 11: Delvis riktige og feilaktige svar med størst frekvens.

Vi ser at det er noen elever som forveksler den sammensatte sannsynligheten (Nr 1) med en slags kvasibetinget sannsynlighet. Denne forvekslingen er omtalt i avsnitt 5.1. Ganske mange elever stipulerer en sannsynlighet (Nr 2) basert på produktet, men andelen synker med undervisning.

Det er overraskende at mange tar den minste sannsynligheten – og at andelen øker med undervisning. Det er mulig at elevene overser den underliggende informasjonen og dermed neglisjerer base rate, omtalt i kapittel 2 som misoppfatning 8C. Andelen som beregner gjennomsnittet av de to hendingene er synkende med undervisning. Jeg har kategorisert dette svaret som feil, men jeg undrer om Diaz og Batanero ville kategorisert det som delvis riktig, siden elevene åpenbart har kjent igjen begge tallene?

Fra resultatdelen husker vi at oppgave 11 og 12 var ganske så like. Blant annet er gjennomsnittscoren i begge oppgavene på 53 prosent. I oppgave 12 er den kvasibetingede sannsynligheten (Nr 1) fremtredende, men synkende med undervisning. Vi ser også at elever uten videre kan droppe den underliggende informasjon (Nr 2), noe som også tilfellet i oppgave 11 (Nr 3).

Kommentar	Nr	svar	Kategori	Før	Etter
Forveksler snitt med «betinget». Ikke produktsetningen.	1	36/91	Delvis	12	5
Forveksling. Overser base rate	2	36 %	Feil	10	9

Tabell. Oppgave 12: Delvis riktige og feilaktige svar med størst frekvens.

I mange tilfeller kan synkrone hendinger gjøres asynkrone. Teorien sier at elever som ikke oppdaget at synkrone hendingene kan betraktes som asynkrone, hadde en tendens til å bruke addisjonssetningen i stedet for produktsetningen. De som hadde fått undervisning i bruk av valgtre brukte produktsetningen og presterte bedre i situasjoner med to asynkrone hendinger (Diaz og Fuente, 2007). I en undersøkelse utført av Fischbein og Gazit (1984) viste det seg at bruken av produktsetningen med tilbakelegg falt enklere enn bruk av produktsetningen uten tilbakelegg (Diaz og Fuente, 2007). Hva sier mine undersøkelser?

Hendingene i oppgave 2, 6a og 7 er asynkrone fordi de er beskrevet slik i teksten. Det samme bør sies om oppgave 11, der teksten åpner opp for at dette er to hendinger som vanskelig skjer samtidig. Oppgave 12 derimot, kan oppfattes som synkron. Likevel er det ingen som har summert sannsynlighetene. Det kan være på grunn av at de ser at sannsynligheten ikke kan være over 1. For øvrig er det fullt mulig å se for seg oppgave 12 som asynkron – først det ene, så det andre. Masteroppgaven har ikke som oppgave å måle forskjellen i synkrone og asynkrone hendinger, men svare på om elevene behersker betinget og sammensatt sannsynlighet, med og uten tilbakelegging i slike settinger - og det må vi konstatere at de gjør. Mine resultater viser følgende funn:

1. **Det var ingen som summerte sannsynlighetene der produktsetningen skulle brukes**
2. **Elevene presterte bedre i bruk av valgtrær og produktsetningen etter undervisning**
3. **Bruken av produktsetningen falt ikke enklere dersom det var tilbakelegg**
4. **Andelen som bruker kvasibetinget sannsynlighet er synkende med undervisning**

Selv vil jeg bestride punkt 3 ovenfor. Min erfaring tilsier at produktsetningen faller enklere dersom vi har en situasjon uten tilbakelegg. Problemet i mine undersøkelser var kanskje at elevene skjønte overgangen mellom oppgave 11 og oppgave 12; bruk produktsetningen! De holdt seg til en bestemt regneregel, omtalt som misoppfatning 6C. Annen misoppfatning som kan knyttes til oppgave 11 og 12 er base rate.

5. **Andelen som gir den minste sannsynligheten som svar er stor, også etter undervisning**

Elevene som kun gir den minste av de to oppgitte sannsynlighetene som svar overser underliggende informasjon. Samlet for begge oppgavene er 14 tilfeller «før» og 16 tilfeller «etter». Dette tilsvarer 11 prosent både før og etter undervisning. Vi kan si at misoppfatninger knyttet til Base Rate holder seg konstant. Samtidig vil jeg si at oppgavene ikke er konstruert for å måle base rate, og at dersom det var prestasjonen knyttet til denne misoppfatningen jeg ville måle, ville jeg også brukt andre typer oppgaver.

5.5 Kombinatoriske problem

Ifølge enhetene fra CPR-testen måler oppgave 8 i hvilken grad elevene kjenner igjen at betingelser begrenser utfallsrommet. I forrige kapittel så vi en økning fra om lag 40 prosent til 80 prosent riktige svar som følge av undervisning. Vi kunne konkludere med at undervisning hjalp med å lage fullstendige utfallsrom og begrense disse etter gitte betingelser. I denne delen drøfter vi større grad misoppfatningene knyttet til oppgavene. Det kan være feil eller delvis riktige svar. Eksempel på delvis riktig svar (Diaz og Batanero, 2009):

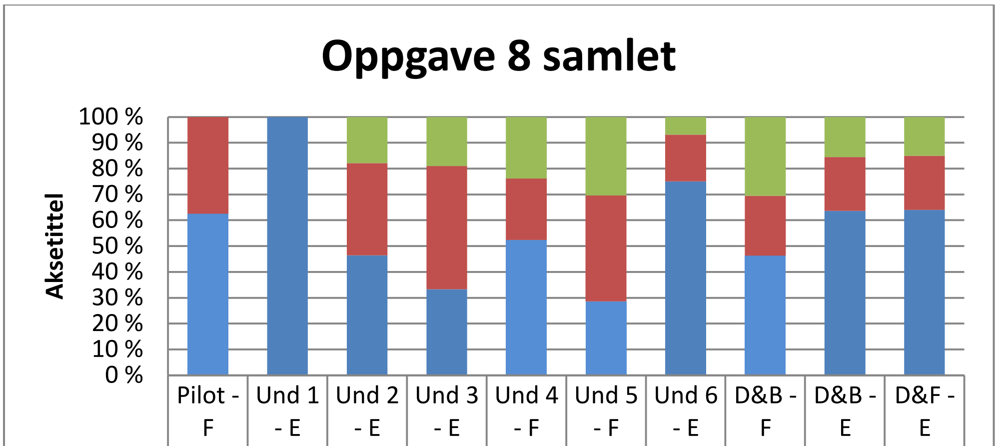
a) $E=\{FFF,\ MFF,\ FMM,\ MMM\}$

b) $E^*=\{FMM,\ MMM\}$

Noen elever klarer ikke å lage et komplett utfallsrom i 8a fordi de utelater noen kombinasjoner mellom gutt og jente, men oppgaven anses som delvis korrekt dersom reduksjonen i oppgave 8b følger det ufullstendige utfallsrommet fra oppgave 8a. Nummering av antall utfall, 8 før og 4 etter betingelse, blir også ansett som delvis riktig. Feilaktige svar får jeg kommentere etter hvert som de dukker opp. I Diaz med flere, var utfallsrommet $E^*=\{MF,\ MF,\ MF\}$ brukt som eksempel på feil svar (Diaz og Batanero, 2009).

> *"The respons is incorrect because the student only considers families with two children instead of families with three, and all the events are identical."*

Ingen av elevene i mine undersøkelser har vært i nærheten av et slikt utfallsrom, og jeg stiller meg undrende til at antallet studenter som har gjort denne feilen hos Diaz med flere, er så stort at de måtte ta med dette eksemplet. Som diskutert i kapittel 4 vil jeg presentere oppgave 8a og 8b samlet. Jeg tok brukte gjennomsnittsverdien av resultatet i begge oppgavene i mine undersøkelser. En nøyere studie av tallene viste at dette var tilnærmet riktig.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E	D&B - F	D&B - E	D&F - E
feil	0,000	0,000	0,179	0,190	0,238	0,304	0,068	0,305	0,155	0,150
delvis	0,375	0,000	0,357	0,476	0,238	0,411	0,182	0,232	0,209	0,210
riktig	0,625	1,000	0,464	0,333	0,524	0,286	0,750	0,463	0,636	0,640

Tabell. Oppgave 8. Andel riktige, delvis riktige og feilaktige svar for alle grupper.

Vi ser at det er en noe større andel delvis riktige (rød) enn det er feil (grønn). Hvilke feil gjør elevene? Tabellen nedenfor prøver å gi svar på dette, og hvordan står det til med de kombinatoriske evnene.

Oppgave 8

Kommentar	Nr	svar	Kategori	Før	Etter
Manglende kombinatorikk i a (4) er riktig redusert til b (2).	1	4 og 2	Delvis	9	14(3)
Mangelfullt eller for stort antall (3, 5 eller 6), men riktig redusert til (2).	2	til 2	Delvis	6	9(2)
Kun antall, ikke skrevet ut kombinasjonene, men riktig redusert.	3	8 til 4	Delvis	1	0(0)
Svært ufullstendig utfallsrom gir ukorrekt reduksjon.	4		Feil	4	3(1)
Andre feil	5		Feil	1	1(0)
Ubesvart	6		Feil	9	3(0)

Tabell. Oppgave 8: Fordeling mellom delvis riktige og feilaktige svar.

Disse tallene må kommenteres. Vi ser at undervisning fører til en kraftig økning av delvise feil, mens resultatet fra kapittel 4 fortalte oss om en forbedring på om lag 40 prosent. Grunnen til de høye tallene i «Etter» gruppen er at de to S1-gruppene i mindre grad var i stand til å lage fullstendige utfallsrom. Ved å fjerne de to S1-gruppene, og bare studere direkte sammenligning (Sam1 og Sam2) får vi et helt annet scenario, se tall i parentesene i tabellen ovenfor. Mine funn er altså at de kombinatoriske problemene blir redusert for vg1T-elevene. For S1-elevene kan vi ikke si om problemet er økende eller minkende med undervisning, men vi kan antyde at problemet i større grad er til stede, selv etter undervisning.

I mine undersøkelser spurte jeg etter bakgrunnsdata for hver elev. Jeg spurte i hvilken grad elevene kjenner til ulike begreper, deriblant utfallsrom. Selv om utfallsrom står nevnt i læreplanen for ungdomsskolen, er det mange elever som før undervisning på vg1T ikke kjenner til begrepet. Bakgrunnsdata av forhåndskunnskaper viser at vel 20 prosent av «før» elevene kjenner til begrepet, mens det tilsvarende tallet for «etter» elevene er på om lag 70 prosent. Dette ble også bekreftet av intervjugruppe 1 (før), der en elev sier:

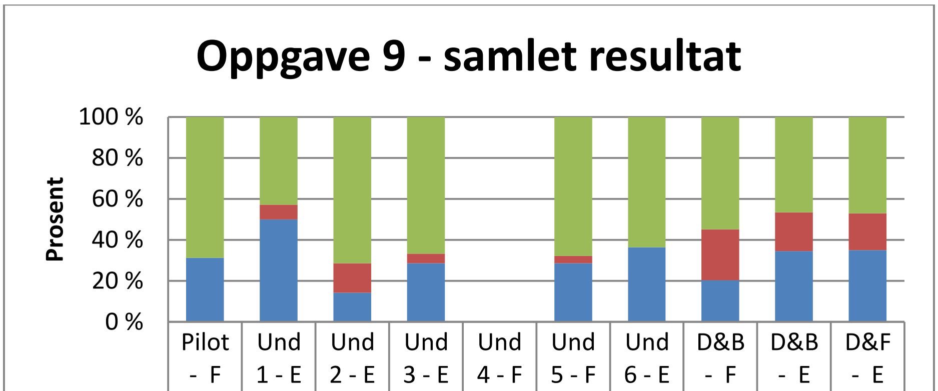
«Hva mener de med utfallsrom? Bare sannsynlighetene for de forskjellige kombinasjonene?»

Etter at elevene i gruppen har funnet det riktige utfallsrommet, utfører de reduksjonen fra betingelsen uten problemer. De to neste intervjugruppene tilhørte gruppene med 100 % score på oppgave 8, det viste seg nytteløst å lokke dem ut på glattisen. I samtale med enkelteleven, viser han som mange andre i undersøkelsene, at selv om utfallsrommet i a ikke er komplett (7 av 8 mulige kombinasjoner), blir svaret i b riktig.

Mange elever tok bare med fire alternativer på 8a og to alternativer på 8b. Av misoppfatningene beskrevet i kapittel 2.7, viste det seg at de fleste feilene i denne oppgaven dreide seg om «Feil ved rekkefølge» - feilen består i forvirringen om rekkefølgen teller eller ikke.

5.6 Misoppfatninger fra oppgave 9

I de neste avsnittene snur jeg retningen for tilnærmingen til misoppfatningene. Jeg starter med oppgavene for så å studere hvilke misoppfatninger som dukker opp. I oppgave 9 vil kunnskapen om å kunne lage komplette utfallsrom gi det riktige svaret 0,5. Dersom elevene overser betingelsen, men identifiserer de to gunstige utfallene og bruker Laplace-regelen riktig (svar: 2/36), kategoriseres dette som delvis riktig. Alle andre svar er kategoriserte som feil, vist med grønn farge i diagrammet nedenfor.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E	D&B - F	D&B - E	D&F - E
Feil	0,688	0,429	0,714	0,667	0,000	0,679	0,636	0,548	0,466	0,47
Delvis	0,000	0,071	0,143	0,048	0,000	0,036	0,000	0,249	0,189	0,18
Riktig	0,313	0,500	0,143	0,286	0,000	0,286	0,364	0,203	0,345	0,35

Tabell. Oppgave 9: Andel riktige, delvis riktige og feilaktige svar for alle grupper.

Vi observerer liten grad av delvis riktige svar, men stor grad av uriktige svar i denne oppgaven. Andelen feilaktige svar er 59 prosent i gjennomsnitt for alle gruppene. Resultat fra kapittel 4 viser at prestasjonen synker etter undervisning. Sammenlignet med oppgave 8, som også dreide seg om begrensning av utfallsrom og der prestasjonen var sterkt økende etter undervisning, er score på oppgave 9 bemerkelsesverdig lav. Gjennomsnittlig score var 57 prosent for alle grupper på oppgave 8 mot 31 prosent på oppgave 9. Oppgaven leder til flere feil og tabellen nedenfor viser hvilke.

Oppgave 9

Kommentar	Nr	svar	Kategori	Før	Etter
Begrenser ikke utfallsrommet. Misoppfatning 7 -Kombinatorisk problem	1	2/36	Delvis	1	4 (1)
Fjernet alle kombinasjoner som inneholder en eller to 6-ere. Produktregelen	2	25/36	Feil	3	10 (5)
Sett på "sum" i stedet for "produkt" av to terninger. System1 - feil	3	0 eller 0 %	Feil	8	16 (9)
Fjernet sekser fra et kast, eventuelt to kast, eller omvendt. Bestemt sannsynlighet	4	1/6, 2/12	Feil	9	3 (3)
Sett på kombinasjonen (6,6). Tatt denne (1/36) eller fjernet den (35/36). Produktregel.	5	1/36, 35/36	Feil	2	2 (1)
Andre feilaktige svar: 1/4, 2/42, 2/10, 1/3, 0.306, 0.2, 10/648, 1/9, 1/9, 3/4, 1/6, 9/16	6		Feil	9	9 (0)
Ubesvart	7		Feil	0	4 (2)

Tabell. Oppgave 9: Fordeling av delvis riktig og ulike typer feil.

Jeg ser igjen at S1-gruppene kan lede til skjevheter mellom elevene som er i «før»- og «etter»-gruppene. Tallene i parentesene ovenfor er de tilsvarende resultatene for «etter» gruppene, uten innblanding fra S1. Uansett observerer vi en mengde med ulike feil og misoppfatninger.
Det er få elever som får delvis riktig. Jeg har ansett dette som et kombinatorisk problem, som er kommentert og diskutert under oppgave 8. Svaret 25/36 er et eksempel på svar som i lys av misoppfatninger kan tolkes i ulike retninger:

- Brukt bestemt sannsynlighet (6A) i en eller annen sammenheng.
- Brukt bestemt regel (6C) i feil sammenheng
- Resonnert etter tilgjengelighet (8A)

Med bestemt sannsynlighet, beskrevet i kapittel 2.6.1, mener vi innarbeidede faste tall som bestemmer sannsynligheten for blant annet kjente forsøk (Devlin, 2014). I oppgave 9 har problemet sitt utspring i et kjent forsøk: terningkast. Terningkast er forbundet med den faste sannsynligheten 1/6. I oppgaven er det mange feilaktige svar med 1/6 eller ulike varianter av dette.

Svarene i nummer 2, 4 og 5 i tabellen ovenfor har direkte tilknytning til sannsynligheten 1/6. Spesielt ser vi utstrakt bruk av produktregelen (nummer 2 og 5), beskrevet under misoppfatning 6C, der denne ikke skal brukes. De kan også ledes til å bruke denne regelen siden det i oppgaven er spurt etter produktet. Svaret kan også være basert på en tilgjengelighetsheuristikk. Vi observerer at misoppfatningene, enten de er forankret i en fast sannsynlighet, fast regneregel eller tilgjengelighet, består etter undervisning. Det mest bemerkelsesverdige i sammenligningen mellom mine undersøkelser og CPR-testen, og en god grunn til at jeg har beholdt kapittel 1.5.4, er feil nummer 3 i tabellen ovenfor. Det er veldig mange elever som ikke leser oppgave 9 godt nok. De leser at det er to terninger, men fordi de anvender system 1, og leser overfladisk, oppdager de ikke at det står «produkt», men løser oppgaven riktig i henhold til ordet «sum», slik de er vant til med de fleste slike oppgaver. Intervjuene vil vise at elevene ville regnet riktig, dersom de hadde anvendt system 2. Feilen med størst frekvens er altså en system 1-feil fra kapittel 1, og ikke en misoppfatning innen sannsynlighet fra kapittel 2. Drøfting av system 1-feil er ikke en del av denne masteroppgaven, men på grunn av svarene i oppgave 9, og ikke minst intervjuene, velger jeg å løfte frem dette.

I intervjugruppe 1 hadde ingen av de tre elevene svart riktig. Det var kanskje ikke overraskende at med en gang den ene eleven sa at svaret var null, hev de to andre seg på. Under forklaringen kom det tydelig fram at eleven tenkte på summen av to terninger i stedet for produktet. Som han sa:

«Det må være null. Det er ingen annen måte å få det (12) på, enn med to seksere».

Jeg påpeker da at det i teksten står produkt. De kommer da raskt fram til (2,6) og omvendt, og like etterpå (3,4) og omvendt. Den samme eleven som var raskt ute med først å si null, svarer nå en halv. Han peker på setningen med betingelsen at produktet skal være lik 12, og han sier:

«Det betyr ikke av alle, men kun av de fire. Og halvparten har ingen sekser»

I intervjugruppe 2 var det også en elev som hadde 0 som svar basert på utfallet (6,6), altså sum. Da påpeker en annen elev at produktet blir 36, og ikke 12. Eleven svarer raskt at hun hadde lest «sum». Etter dette blir de raskt enige om det riktige svaret.
I det semistrukturerte intervjuet sier eleven først at svaret er «Null» med forklaringen:

«Terningen viser 6, og det er to terninger som er like, så 6 ganger 2 blir 12.»

Han vingler litt mellom svarene 0 og 1, enten – eller. Konfrontert med det riktige svaret og spørsmålet om hvor mange måter et produkt av to terninger kan bli 12 på, svarer han:

«På én måte, 6 pluss 6 er lik 12.»

Mitt neste spørsmål om hva et produkt er, fremskaffer øyeblikkelig litt latter hos kandidaten. Han sier seg enig og bytter til alternativet 0,5. Alle intervjuene bekreftet altså system 1-feilen, men også at de ville beregnet sannsynligheten riktig, dersom de hadde anvendt system 2. Intervjugruppe 3 diskuterte ikke oppgave 9. Intervjuene viser at mange elever ikke kobler inn system 2, men bare forholder seg til system 1. Dette kan ha sin forklaring i at forsøk med kast av to terninger i norske oppgaver, ofte er formulert med sum som utgangspunkt, sjelden produkt. Min konklusjon er at system 1-feil hindrer riktig løsning av oppgaven. Jeg har valgt å la «conjuction fallacy» stå alene som en misoppfatning knyttet til oppgave 3, men det er overveiende sannsynlig at dersom elevene ble bedt om å ta seg god tid til å lese oppgaven nøye, ville prestasjonen også vært bedre fordi vi da hadde redusert muligheten for system 1-feil. På den ene siden ville jeg før undersøkelsene startet sagt at system 1-feil på oppgave 3 er mindre overraskende enn samme feil på oppgave 9 – og det er kanskje det som er tilfellet også, siden prestasjonen på oppgave 3 er vesentlig lavere enn på oppgave 9. På den andre side fikk jeg sterkere bekreftelser i intervjuene på system 1-feil i oppgave 9. Egentlig har alle oppgaver en større eller mindre grad av system 1 tenkning, og derfor mulighet for en system 1-feil.

5.7 Misoppfatninger fra oppgave 10

I kapittel 2 sa jeg at heuristikker er universelle; de kan dukke opp hvor som helst og hos hvem som helst – når som helst. Jeg har bare skrevet om et fåtall heuristikker, men finner vi dem igjen i svarene til elevene?

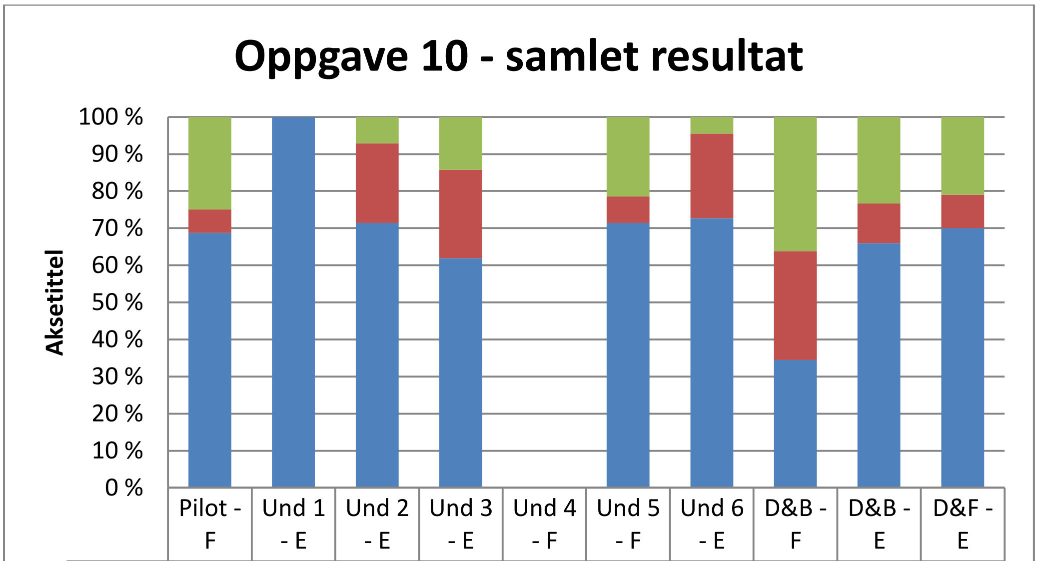
Heuristisk representativitet vil si å vurdere sannsynligheten for en hending ut fra hva som tidligere har hendt. Effekten av siste observasjoner, «recency effects», deles inn i to retninger:

- Positiv – forventer samme utfall etter en lang serie (trenden holder seg)
- Negativ – forventer motsatt utfall etter en lang serie (trenden snur)

I oppgave 10 er spørsmålet sannsynligheten for oddetall i et terningkast etter en serie med 10 oddetall og 5 partall. Hva som anses som delvis riktige og feilaktige svar i henhold til Diaz og Batenero sin kategorisering, gir en beskrivelse av de ulike misoppfatningene fra kapittel 2.5.

I oppgave 10 skal elevene beregne betinget sannsynlighet I et forsøk med tilbakelegging, altså uavhengighet. Antall øyne som terningen viser er uavhengig av tidligere resultat og sannsynligheten for oddetall er derfor 0,5. En estimert verdi, 2/3, basert på frekvenser til tidligere resultater blir ansett som delvis riktig svar fordi eleven ikke har sett på terningen som rettferdig, selv om dette står i oppgaveteksten. Det er feil løsning å resonnere i henhold til «Gambler's fallacy» (trenden snur) og tro at sjansen for oddetall i neste forsøk er 5/15. Siden 10 oddetall av 15 mulige ikke oppfattes som et representativt utfall for en rettferdig terning, innvirker elevenes ønske om balanse på en slik måte at de holder partall som mest sannsynlig ved neste korsvei. Det vil gi en bedre representasjon av en rettferdig terning.

I den første undersøkelsen til Diaz og Fuente viste mange studenter at de behersket å løse problemer med i betinget sannsynlighet med tilbakelegg. Hele 70 % svarte riktig på denne oppgaven (Diaz og Fuente, 2007). I forsøk gjennomført av Green (1982) og Konold med flere (1993), trekker de konklusjonen at majoriteten av barn i 12-årsalderen i stor grad svarer riktig på et lignende spørsmål. Spørsmålet dreier seg om myntkast, og er ikke helt likt, men de fleste gjenkjenner uavhengighet i oppgaven (Diaz og Fuente, 2007). Tabellen nedenfor viser andelen av riktige, delvis riktige og feilaktige svar.



	Pilot - F	Und 1 - E	Und 2 - E	Und 3 - E	Und 4 - F	Und 5 - F	Und 6 - E	D&B - F	D&B - E	D&F - E
Feil	0,250	0,000	0,071	0,143	0,000	0,214	0,045	0,362	0,233	0,210
Delvis	0,063	0,000	0,214	0,238	0,000	0,071	0,227	0,294	0,107	0,090
Riktig	0,688	1,000	0,714	0,619	0,000	0,714	0,727	0,345	0,660	0,700

Tabell. Oppgave 10: Andel riktige, delvis riktige og feilaktige svar.

Vi ser at det er marginalt med feilaktige svar blant elevene i mine undersøkelser, og noe større for Diaz med flere sin del. Det er også en tendens til at andelen delvis riktige svar i større grad blir redusert med undervisning enn andelen feilaktige svar. Andre svar som jeg har kategorisert som delvis riktige svar er 10/16 og 11/16. Begge svarene er tilnærmet lik 2/3, kategorisert som delvis riktig, men samtidig er svaret 10/16 en helning mot Gamblers Fallacy, fordi denne eleven antar at (16. kast) det er på tide med et partall. Jeg ville utfra Diaz og Batanero sin vurdering muligens valgt 10/16 som feil svar og 11/16 som delvis riktig. Samtidig kan vi argumentere for at begge svar er feilaktige. Fordeling etter misoppfatninger i kapittel 2.5 vises i tabellen nedenfor.

Oppgave 10

Kommentar	Nr	svar	Kategori	Før	Etter
Representativitet. Recency effects - positiv. Trenden holder seg.	1	2/3	Delvis	3	11 (5)
Representativitet. Recency effects - negativ. Gamblers fallacy.	2	1/3	Feil	5	2 (0)
Sannsynligheten for et enkelt utfall. Bestemt sannsynlighet (6A)	3	1/6	Feil	3	2 (0)
ubesvart, (eller 100 %)	4		Feil	2	3 (1)

Tabell. Oppgave 10: Fordeling av heuristisk representativitet.

Vi ser en kraftig økning i troen på at trenden vil holde seg. Selv om 6 elever av de 11 som svarer 2/3 kommer fra S1-gruppene, viser tallene en økning fra «Før» til «Etter». For Gambler's fallacy får vi motsatt effekt. Undervisning virker positivt på misoppfatningen «Gambler's fallacy», men negativt på troen om at trenden holder seg. Hva som er positivt og negativt her, er ikke lett å si. Tallenes tale blir

uklar fordi jeg i mine undersøkelser strever med for høy score før undervisning. Skal jeg peke på et funn må det være at elevene flytter seg fra den ene veikanten til den andre. De forkaster negativ recency til fordel for den positive. Hvis det er noe positivt i det, må det være at positiv recency i CPR er kategorisert som delvis korrekt. Et lignende problem med lignende heuristikker, men med andre sannsynligheter, er skissert av Tversky og Kahneman (side 303 i Tversky og Kehneman, 1983). Som vi observerte i oppgave 9 er også misoppfatningen om at det til et bestemt forsøk er knyttet en bestemt sannsynlighet til stede. Svaret 1/6 for et terningkast i oppgave 10 forekommer både før og etter undervisning og kan være basert på heuristikken om fast sannsynlighet, «anchoring». Svaret kan også være basert på tilgjengelighetsheuristikk som beskrevet i kapittel 2.8.1. Sannsynligheten for en hending som er basert på hvor lett noe kommer frem som løsning i sin egen tenkning, kaller vi tilgjengelighetsheuristikk. Oppgave 9 (to terninger) og oppgave 10 (en terning) er to oppgaver der elever kan resonnere i henhold til denne heuristikken, men jeg kan vanskelig konkludere om dette.
I utfallstilnærming, «outcome approach» (kapittel 2.8.2), er hovedmålet i sannsynlighet å forutsi utfallet i et enkelt forsøk. En av elevene svarte 100 % sannsynlig for oddetall (Nr 4). Resonnerer eleven da ifølge utfallstilnærming, «outcome approach», eller er det bare ren tilfeldighet? Her er det vanskelig å få bekreftet hva som er hva, til det er tilfellene for få – og her bare ett. I slike tilfeller kan vi se viktigheten av et intervju – og at blandet metode kan avsløre flere ting som vi ellers ikke hadde fått svar på. Selv om jeg i dette tilfellet ikke fikk svar, kunne jeg med bedre organisering av intervjuene kanskje fått svar. Nå skal det sies at intervjuene ikke tilførte noe ekstra informasjon til oppgave 10. Alle elevene var skråsikre på at det riktige svaret var 0,5, og var ikke interessert i noen videre diskusjon om dette. Eleven i det semistrukturerte intervjuet bekreftet skråsikkerheten ved å si at han ikke hadde kastet så mye som et blikk på de tidligere resultatene når han i oppgaven leste ordene «terning» og «rettferdig». Min konklusjon er at vi på oppgave 10 finner igjen noen heuristikker slik de er beskrevet i teorien, men at undervisning reduserer andelen feil som kan skyldes heuristikker. Jeg baserer den siste konklusjonen på tall uten S1-gruppene. For andre heuristikker er det vanskelig å konkludere.

5.8 Andre misoppfatninger fra kapittel 2

Misoppfatningen om lik sannsynlighet (fra kapittel 2.6.2) kan vi finne i oppgave 2, 3, 4, 5 og 6.
I oppgave 2 har alle som ikke svarte riktig, unntatt én elev, krysset av for «like sannsynlig». Har de da en forståelse av at utfallene deler sannsynligheten likt mellom seg, som beskrevet i kapittel 2.6.2 6B Lik sannsynlighet, eller har de ikke oppfattet at situasjonen er uten tilbakelegg? Jeg heller mot det siste og får noe bekreftelse fra intervjusituasjonen. En av elevene blir konfrontert med sitt gale svar; at begge alternativene i og ii er like sannsynlig. Da svarer hun:

«Eh, ja. Jeg tror ikke at jeg tenkte riktig med tilbakelegging og sånt.»

Jeg forbeholder oppgave 3 til kun å besvare conjunction fallacy og oppgave 4 til kun å besvare misoppfatningen, eller forvekslingen, mellom to betingede sannsynligheter. Da gjenstår oppgave 5 og 6. I oppgave 5, som måler forståelsen om inversjon av tidsaksen, svarer 71 prosent av elevene i mine undersøkelser ½, altså lik sannsynlighet. Jeg synes likevel at det er vanskelig å bruke denne oppgaven til å svare på misoppfatningen om lik sannsynlighet. Når elevene først ikke klarer inversjon av tidsaksen, altså at de overser betingelsen og beveger seg fysisk tilbake til de to kanalene, er det riktig at sannsynligheten er lik for kanal 1 og kanal 2. Da gjenstår kun oppgave 6. Oppgave 6b kan vanskelig brukes, med samme argumentasjon som for oppgave 5. I oppgave 6a svarer kun 3 prosent av mine elever at sannsynligheten er lik ½, men dette svaret kan, som for oppgave 2, være et resultat av at elevene har oversett om det er med eller uten tilbakelegging. Min konklusjon er at testen, i motsetning til hva jeg trodde på forhånd, ikke kan brukes til måle misoppfatningen om lik sannsynlighet. I oppgave 6 er det to kuler av hver farge. Det første målet på sannsynlighet som trolig dukker opp hos de fleste er 0,5 – anvender tilgjengelighetsheuristikk. På bakgrunn av drøftingen mellom oppgave 6a og 6b som nevnt ovenfor, utelukker jeg tilgjengelighetsheuristikken som mulig misoppfatning i denne oppgaven. Ikke fordi det ikke kan være det, men fordi testen ikke kan konkludere med at så er tilfelle. Jeg konkluderer med at testen og denne masteroppgaven ikke gir svar på alle misoppfatninger som er omtalt i kapittel 2.

6 Oppsummering, Konklusjon og Implikasjon

Dette kapitlet samler trådene fra kapittel 4 og 5 og viser til vesentlige funn som igjen leder til konklusjoner. Med konklusjonene som bakgrunn drøfter jeg implikasjoner for undervisning i sannsynlighetsregning.

6.1 Oppsummering og 12 vesentlige funn

Det var ingen spesiell grunn til at det ble akkurat 12 funn, og egentlig ender det opp med 13, men det siste får vente. Vi får også vente med å se i hvilken grad de 12 funnene kan få betegnelsen «vesentlige». Hvem er det i så fall vesentlig for?

Kontekst kan lede til forvekslinger ved beregning av ulike sannsynligheter. Oppgave 1 dreide som om å lese/beregne sannsynligheter fra en krysstabell. Resultat fra kapittel 4 viste at undervisning hjelper for å unngå slike forvekslinger - den samlede prestasjonen ble bedre. Vi har funn nummer 1:

1 **Elevene behersker oppgaver med kontekst bedre etter undervisning**

Dette er et viktig punkt siden mange oppgaver i sannsynlighetsregning har og må ha kontekst. Drøftingen av oppgave 1 viste at andelen elever som forvekslet ulike sammensatte sannsynligheter holdt seg konstant mellom «før»- og «etter»-gruppene. Det betyr at forbedringen i funn 1 skyldes en reduksjon av «andre feil». Av de ulike sammensatte sannsynlighetene som er omtalt finner vi spesielt en negativ utvikling i resultatet på oppgave 4. Denne oppgaven målte forvekslingen mellom to motsatt betingede sannsynligheter. Vi har funn nummer 2:

2 **Elevene har vansker med å se forskjell på to motsatte betingede hendinger**

Den negative utviklingen var marginal, men utgangspunktet var også lavt. Funn 1 peker på en forbedring rundt ulike sannsynligheter i oppgaver med kontekst, mens funn 2 viser at forbedringen ikke skjer på alle områder.

Snittet mellom to hendinger kan ikke være større enn den enkelte hending. Resultatet på oppgave 3 viste en forbedring i prestasjonen med undervisning, men forbedringen er liten, og det samlede resultat for alle grupper på 28 prosent gir oss funn nummer 3:

3 **Misoppfatningen conjunction fallacy står sterkt – før og etter undervisning.**

Intervjuene fra drøftingen bekrefter misoppfatningen og at elevene opplever dette som vanskelig.

Oppgave 5 og 6b dreide seg om inversjon av tidsaksen. Resultatet viste at prestasjonen til elevene sank etter undervisning. Intervjuene fra drøftingen i etterkant bekreftet at elevene har store problemer med å tilbakedatere effekten av en årsak. Mine funn viser at

4 **Evnen til å beregne betinget sannsynlighet med omvendt tidsakse er lav**

5 **Misoppfatningen om omvendt tidsakse står sterkt og er økende med undervisning**

I intervjusituasjonene var det disse oppgavene som skapte størst frustrasjon blant elevene.

I oppgaver med suksessive forsøk, med og uten tilbakelegging og bruk av produktsetningen presterte elevene godt. Mine funn viser at

6 **Elevene har en intuitiv forståelse av produktsetningen**

7 **Prestasjoner i problemer som er kompatible med tidsaksen øker med undervisning**

Det siste punktet virker positivt inn på funn 1; elevene presterer bedre etter undervisning. Funn 6 viser til elevenes intuitive forståelse av produktsetningen. Dette funnet er ikke entydig i mine undersøkelser. Oppgave 11 og 12 krever korrekt bruk av produktsetningen og resultatet viste at Und5 fikk lav score på disse to oppgavene (32,1 % på begge).

Elevenes kombinatoriske evner er viktige. Mine undersøkelser gir funn 8

8 Evnen til å utarbeide fullstendige utfallsrom er økende med undervisning

Som et resultat av dette presterer elevene også bedre på oppgave 8. Dette gir oss funn 9:

9 Evnen til å redusere utfallsrommet ut fra gitte betingelser er økende med undervisning

Selv om resultatet i oppgave 8 viser en klar forbedring i å redusere utfallsrommet, viser resultatene i oppgave 9 at dette nødvendigvis ikke gjelder alle typer oppgaver. Jeg står ved påstanden ovenfor, men oppgave 9 viser ikke tilsvarende forbedring i resultat som oppgave 8. Oppgave 9 viste i mine undersøkelser at system 1-feil fort kan oppstå. Det var mange elever som leste «sum» i stedet for «produkt». Dette gir funn 10:

10 Mange elever gjorde system 1-feil på oppgave 9

Denne type feil kan ha gått igjen på andre oppgaver også, men på oppgave 9 er det gjennom intervjuene verifisert at elevene gjør en slik feil.

Heuristikker dukker opp hos elever også etter undervisning. I oppgave 10 kunne jeg observere effekten av siste observasjoner – tilbakeblikk. Vi finner funn som:

11 Andelen som resonnerte etter effekten av siste observasjoner holdt seg konstant

12 Elevene flyttet seg fra negativ (Gambler's fallacy) mot positiv (trenden holder seg)

Effekten av siste observasjoner heter recency effects i engelsk litteratur. Jeg fikk ingen støtte i intervjuene for disse funnene og må stole på den kvantitative delen av undersøkelsen.

Hvordan samsvarer mine funn med Diaz og Batanero og gangbar teori? Mine 12 funn i listen nedenfor er i større grad kronologisk enn de er vurdert ut i fra hvor vesentlige de er. De tre siste funnene er ikke omtalt i undersøkelsene til Diaz med flere (Diaz og Batanero, 2009). Mine funn er relativt konsistente med andre internasjonale funn og bekrefter at de samme misoppfatningene eksisterer på tvers av landegrenser.

12 Funn	
1	**Elevene behersker oppgaver med kontekst bedre etter undervisning** Funnet samsvarer med teori og bekrefter resultat fra Diaz og Batanero.
2	**Elevene har vansker med å se forskjell på to motsatte betingede hendinger** Funnet samsvarer med teori og bekrefter resultat fra Diaz og Batanero.
3	**Misoppfatningen conjunction fallacy, står sterkt – før og etter undervisning** Funnet samsvarer med teori og bekrefter resultat fra Diaz og Batanero.
4	**Evnen til å beregne betinget sannsynlighet med omvendt tidsakse er lav** Funnet samsvarer med teori og bekrefter resultat fra Diaz og Batanero.
5	**Misoppfatningen omvendt tidsakse står sterkt og er økende med undervisning** Funnet er konsistent i mine undersøkelser fordi jeg får negativt resultat på begge oppgavene som måler denne misoppfatningen. For Diaz og Batanero viser det seg at misoppfatningen reduseres i oppgave 5, men er kraftig økende i oppgave 6b. Jeg finner ikke annen teori som sier at denne misoppfatningen er økende med undervisning (sjekk).
6	**Elevene har en intuitiv forståelse av produktsetningen** Mine resultater er samsvarende med teori. Elevene i mine undersøkelser viser en større grad av forståelse enn Diaz og Batanero.
7	**Prestasjoner i problemer som er kompatible med tidsaksen øker med undervisning** Mine resultater er samsvarende med teori. Elevene i mine undersøkelser viser en større grad av forståelse av slike problemer enn Diaz og Batanero.
8	**Evnen til å utarbeide fullstendige utfallsrom er økende med undervisning** Funnet samsvarer med teori og bekrefter resultat fra Diaz og Batanero.
9	**Evnen til å redusere utfallsrommet ut fra gitte betingelser er økende med undervisning** Funnet samsvarer med teori og bekrefter resultat fra Diaz og Batanero.
10	**Mange elever gjorde system 1-feil på oppgave 9** Vi kjenner igjen dette funnet fra teorien, men det er ikke omtalt hos Diaz og Batanero. Jeg synes at system 1-feilen ble så fremtredende i svarene på denne oppgaven at jeg tok det med.
11	**Andelen som resonnerte etter effekten av siste observasjoner holdt seg konstant** Dette er et funn hos meg, men er ikke omtalt hos Diaz og Batanero. Jeg har ikke undersøkt om funnet er konsistent med andre undersøkelser.
12	**Elevene flyttet seg fra negativ (Gambler's fallacy) mot positiv (trenden holder seg)** Funnet kommer som et underpunkt til funn 11 og er på siden av masteroppgaven. Jeg finner derfor ikke tilsvarende funn hos Diaz og Batanero og jeg har heller ikke leitet etter slike funn i annen teori. Jeg ser liten eller ingen logisk forklaring, annet enn at elevene etter undervisning i større grad er i stand til å lese frekvenser og dermed i mindre grad er følelsesstyrt – slik gamblere ofte er.

Funn 11 og 12 ovenfor relaterer seg til representasjonsheuristikk. Det er ikke en sentral del av denne masteroppgaven, men når jeg ser forflytningen fra den ene veikanten (negativ) til den andre (positiv), griper jeg sjansen til å studere tallene nærmere. Mitt spørsmål er om det er de samme elevene som forflytter seg fra den ene siden til den andre eller om dette er helt tilfeldig? For å svare på dette må jeg se på de undersøkelsene fra Und5 og Und6 som har de samme elevene i en «før»- og «etter»-gruppe. Jeg beholdt kandidatnummeret i tilfelle jeg skulle sammenligne svarene fra samme elev.

- Det er kun to elever som har svart i henhold til Gamblers' fallacy (1/3) før undervisning. Begge disse elevene har byttet side til (2/3), posistiv recency, etter undervisning.
- En elev har positiv recency (2/3) både før og etter.
- To elever har gått fra riktig svar (1/2) til positiv recency (2/3) og én elev har gjort motsatt.

Det øverste punktet er litt interessant. Hvorfor har elevene byttet ståsted? Dessverre oppdaget jeg dette byttet for sent til å få svar gjennom intervjuene og tallmaterialet er for lite til at jeg tør si noe spesielt eller konkludere. Noe for en annen masteroppgave?

6.2 Konklusjon

I denne masteroppgaven har jeg sett nærmere på misoppfatninger innen betinget sannsynlighet i den videregående skole. Jeg har gjennom CPR-testen, eller et utvalg av oppgavene fra denne, og gjennom intervju påvist endringer i prestasjoner innen betinget sannsynlighet og samtidig påvist ulike misoppfatninger, enten de er knyttet til betinget sannsynlighet eller ikke.

6.2.1 Svar på problemstillingen

I hvilken grad har jeg fått svar på problemstillingen?

Hvilken innvirkning har undervisning i skolen på misoppfatninger innen betinget sannsynlighet?

Vil undervisning redusere antall misoppfatninger?

Hvilke misoppfatninger består, øker eller minker etter undervisning?

Undervisning hjelper elevene til å forstå konteksten og identifisere betingelser. Dette gjør elevene i bedre stand til å løse problemer med betinget sannsynlighet. Mine undersøkelser viste en forbedring i prestasjonene før og etter undervisning på

6 % for alle gruppene (inklusive S1 og Und4)

10 % for direkte sammenligning (Sam1 og Sam2)

Til sammenligning var forbedringen for Diaz med flere på hele 19 %, nesten dobbelt så bra som mine utvalgte grupper. Jeg gjentar at elevene i mine undersøkelser ofte hadde et bedre utgangspunkt før undervisning enn studentene til Diaz med flere. Dette førte naturligvis til mindre utslag på forskjellene i prestasjonene før og etter undervisning. Både 10 % og 19 % forbedring vil jeg si er en markant fremgang. Som en direkte konsekvens av at elevene presterer bedre, reduseres antall feil som elevene gjør. Jeg skiller feil fra misoppfatninger slik disse er beskrevet i kapittel 2. For å svare på om undervisning reduserer antall misoppfatninger så måtte jeg stikke dypere inn i materien. For eksempel kan jeg vise til oppgave 1, der antallet feil er redusert til det halve, mens den omtalte misoppfatningen ikke er redusert. For andre oppgaver kan misoppfatningene indirekte leses ut fra prestasjonene til elevene; prestasjonen forbedres -> misoppfatningen reduseres – eller omvendt prestasjonsbilde. Det er fullt mulig å vise til en forbedring og et dårlig resultat. Med dette mener jeg at en misoppfatning kan reduseres etter undervisning, men likevel bestå - og det i stor grad. Et eksempel på dette er «conjunction fallacy» som i mine undersøkelser og med direkte sammenligning ble redusert med hele 11 %, men jeg sier samtidig at misoppfatningen består, fordi økningen i andelen riktige svar gikk fra 25 % til 36 %. Etter undervisning er det altså vel 1/3 av elevene som resonnerer riktig. Med samme argumentasjon kan vi si at misoppfatningen med omvendt tidsakse består, og faktisk er økende etter undervisning. Jeg mistenker at en undervisning som støtter seg på beregning med to hendinger og bruk av valgtrær og som mer eller mindre alltid er kompatible med tidsaksen, har direkte negativ innvirkning på beregninger med to hendinger og omvendt tidsakse. Forvekslingen mellom to motsatte betingede sannsynligheter består og holder seg konstant. I det store og det hele bekrefter mine funn i denne masteroppgaven funnene til Batanero og Diaz.

- Undervisning hjelper elever til å forstå, identifisere og løse problemer med betinget sannsynlighet
- Alle misoppfatninger omtalt i kapittel 2 vedrørende to omvendt betingede sannsynligheter faller vanskelig, og er til noe grad økende med undervisning
- Misoppfatningen om at snittet er større enn den enkelte hending står sterkt
- Undervisning hjelper på evnen til å begrense utfallsrommet etter gitte betingelser

Det siste punktet er vesentlig når det gjelder betinget sannsynlighet. Klarer elevene å begrense utfallsrommet etter gitte betingelser, viser det at undervisning hjelper. Mange av misoppfatningene beskrevet i kapittel 2 hadde bred forekomst, og mange elever gjorde ingen forbedring på dette, selv med undervisning. Prosentandelen av studentene til Batanero og Diaz som gjorde conjuction fallacy, forvekslet to motsatte betingede sannsynligheter og feilslutningen om tidsaksen var like stor før og etter undervisning (Diaz og Batanero, 2009). For mine grupper var det tilsvarende resultatet også nedslående for disse misoppfatningene.

Det er mange svar i undersøkelsene som kan tolkes i lys av mange ulike misoppfatninger. For å få klarhet om elevene mener den ene eller den andre misoppfatningen, må ulike tiltak gjøres:

- Bruke andre oppgaver eller komplettere for å se et mønster
- Bruke intervju for å få et sikrere grep om hvilken misoppfatning en elev kan ha hatt

Testen kan svare på mye, men ikke alt. Testen svarer best på evnen til å gjøre riktige beregninger, avsløre misoppfatningene knyttet til forveksling mellom to omvendt betingede sannsynligheter, hendinger med omvendt tidsakse, conjunction fallacy og evnen til å begrense utfallsrommet i henhold til gitte betingelser. Testen er lang og kan virke utmattende fordi det i tillegg er mange ukjente oppgaver for elevene. Dette kan lede til system 1-feil, noe som kan gi uriktige utslag. Jeg er ganske overbevist om at vi ville fått bedre score på «senere» oppgaver som 11 og 12 dersom disse var plassert fremst, men samtidig må det sies at det ligger en grundig tenkning bak rekkefølgen til oppgavene. Det er for eksempel en naturlig rekkefølge i oppgave 1, 2, 7 og til slutt 11 og 12, ved en gradvis utvidelse av det samme emnet.

6.2.2 Andre forsøk som viser at undervisning hjelper

Som et bevis på at instruksjon kan påvirke elevers forståelse på et tidlig stadium, gjennomførte Jones et al (1999) en instruksjonsperiode som viste en signifikant forbedring av elevers prestasjoner om begreper innen sannsynlighet. Resultatene fra undersøkelsen viste også at god forståelse av begreper som utfallsrom og teoretisk sannsynlighet for en hending, er nødvendig for å kunne utvikle forståelsen av betinget sannsynlighet (side 228 i Tarr og Lannin, 2005).

Videre bevis for undervisnings påvirkning kan finnes hos Tarr (1997). Undersøkelsen viste signifikante forbedringer i betinget sannsynlighet. Spesielt i betinget sannsynlighet der 19 av 26 elever ble plassert på nivå 1 og 2 før undervisning, mens etter undervisning var 22 av de 26 elevene plassert på nivå 3 og 4. (5. klasse). Tilsvarende forbedringer ble også funnet i uavhengige hendinger. (side 228 i Tarr og Lannin, 2005). Siden dette gjelder elever på 5. trinn, er beskrivelse av de fire ulike nivåene utelatt både her og i appendiks. Nivå 1 er lavest og nivå 4 er høyest.

Tabellen nedenfor gir en kort oppsummering av hver misoppfatning fra kapittel 2:

Nr	Misoppfatning fra kapittel 2
1	**Kontekst og forveksling** Etter undervisning blir elevene flinkere til å lese oppgavene riktig. De er i større grad i stand til å skille mellom de sammensatte sannsynlighetene, og dermed redusere antall forvekslinger.
2	**Snittet mellom to hendinger, conjunction fallacy** Elevene blir flinkere med undervisning, men fortsatt er graden av riktige svar lav.
3A	**Motsatte betingede sannsynligheter - forveksling og likhet** Mange elever har den misoppfatningen at det ikke er forskjell på to motsatte betingede sannsynligheter. Lav grad av riktige svar også etter undervisning.
3B	**Motsatte betingede sannsynligheter - inversjon av tidsaksen** Elevene har store vansker med å godta at betingelser tilbakedateres. Misoppfatningen med omvendt tidsakse står sterkt. Undervisning har negativ innvirkning. Lav grad av riktige svar.
3C	**Motsatte betingede sannsynligheter - bayesiske problemer** Oppgaven som målte 3C falt ikke heldig ut i mine undersøkelser. Jeg fikk rett og slett ikke svar.
4	**Synkrone og asynkrone hendinger** Det ble ikke målt i hvilken grad det var forskjell i antall misoppfatninger mellom synkrone og asynkrone hendinger, men på beregninger av ulike sannsynligheter med og uten betingelser. Mine undersøkelser viste at elevene har en god intuitiv forståelse av betinget sannsynlighet og andre sammensatte sannsynligheter som er kompatible med tidsaksen. Elevene blir flinkere til å anvende produktsetningen riktig etter undervisning.
5	**Heuristikker, siste observasjoner og representativitet** I hvor stor grad elevene støtter seg til heuristikker er varierende. Den mest fremtredende heuristikken er at elevene støtter seg til de siste observasjoner i uavhengige forsøk. Mine funn viser at elevene før undervisning i større grad tror at trenden skal snu, mens de etter undervisningen i større grad heller til at trenden holder seg.
6	**Den mest vanlige, bestemt sannsynlighet, lik sannsynlighet, bestemt regel** Det ble lansert tre kandidater til «den mest vanlige misoppfatningen», og det var noen svar i noen oppgaver som kunne tolkes i retning av disse misoppfatningene. Samtidig viste andre svar at det var vanskelig å konkludere med hvilken misoppfatning hver enkelt elev hadde gjort. Testen gav ikke svar på om de feilaktige svarene skyldtes det ene eller det andre.
7	**Kombinatoriske problem** Mine undersøkelser viser at elevene er betraktelig bedre i kombinatorikk og evnen til å redusere utfallsrommet etter gitte betingelser etter undervisning enn før undervisning.
8	**Heuristikker, tilgjengelige utfall, utfallstilnærming, base rate** Dette er misoppfatninger som jeg trodde skulle dukke opp, men som ikke gjorde det, eller gjorde det i mindre grad. I noen tilfeller kan vi påstå at elevene neglisjerte base rate, men oppgavene var ikke designet for å avsløre eventuell manglende hensyn til base rate, og derfor ble det også vanskelig å konkludere med at så var tilfelle. Et svar kunne muligens tolkes etter utfallstilnærming, men uten intervju er det vanskelig å konkludere. Det samme gjelder for svar som kunne tolkes som et resonnement i lys av tilgjengeligheten til et utfall.
Syst1	**System 1-feil. Beskrevet i kapittel 1** Denne feilen var nok til stede i flere oppgaver enn i oppgave 9, men i denne oppgaven fikk jeg en solid bekreftelse gjennom intervju.

Tabell: Oppsummering misoppfatninger. Nummer 6 og 8 hadde flere misoppfatninger som er slått sammen i denne konklusjonen.

6.3 Implikasjoner og veien videre

På bakgrunn av mine funn, resultat fra CPR-testen til Diaz og Batanero og gangbar teori vil jeg komme med noen implikasjoner til drøfting. Hva bør vektlegges for å styrke og forbedre både det som er bra og det som ikke fungerer like bra? Jeg vil se nærmere på noen mulige tiltak innen

- Læreplanen – inn eller ut og opp eller ned?
- Utdanning – formell og uformell kompetanse?
- Undervisning – kan og bør noe endres?

Hva kan vi gjøre for at elevene skal lære sannsynlighetsregning på en god og sunn måte og at de ikke skal synes at sannsynlighetsregning og statistikk er så vanskelig?

6.3.1 Læreplanen

Det er alltid store diskusjoner rundt hver revisjon av læreplaner i matematikk. Er punktene i læreplanene plassert på riktig nivå? Jeg vil skille mellom følgende nivåer i læreplanene:

- Læreplanen for 8. – 10. trinn
- Læreplanen for vg1 – 1T
- Læreplanen for vg2 – R1/S1

Jeg vil ta utgangspunkt i funnene i min oppgave og mine egne erfaringer. Først vil jeg nevne at jeg er av den oppfatning at mange emner og begreper innen sannsynlighet kan forskyves nedover i systemet, fra vg2 til vg1 og fra vg1 til ungdomsskolen. Hva som skal ut av læreplanen for å få noe inn, diskuteres i mindre grad. Her kommer en liste – nødvendigvis ikke prioritert.

Kunnskap om utfallsrom er en essensiell del av å kunne finne riktig løsning til et problem knyttet til sannsynlighet. Å kunne utarbeide et fullstendig utfallsrom ses på en av de fire viktigste komponentene for å kunne beherske sannsynlighetsregning (side 12 i Brynat og Nunes, 2012). Utdrag fra læreplanen for ungdomsskolen viser at begrepene er nevnt spesielt.

- **beskrive utfallsrom og uttrykkje sannsyn som brøk, prosent og desimaltal**
- **drøfte og løyse enkle kombinatoriske problem**

Mine funn viste at det nytter å arbeide med dette

8 Evnen til å utarbeide fullstendige utfallsrom er økende med undervisning

9 Evnen til å redusere utfallsrommet ut fra gitte betingelser er økende med undervisning

Etter min mening burde det i ungdomsskolen blitt lagt enda større vekt på begrepet utfallsrom og det kombinatoriske arbeidet med å lage slike. Nå har jeg ikke inngående kjennskap til omfanget av det arbeidet som gjøres med nettopp kombinatorikk i ungdomsskolen, men funn – og her kommer funn nummer 13 – basert på bakgrunnsdata fra undersøkelsene mine viser at jeg kan ha rett i min påstand:

- Av de elevene som ikke har hatt undervisning på vg1 svarer kun 23 % at de har kjennskap til begrepet utfallsrom.

Funn nummer 13 er altså den overraskende lave andelen som har kjennskap til begrepet utfallsrom, selv om dette begrepet er tydelig i læreplanen for ungdomsskolen, men det fortsetter:

- Av de elevene som har hatt undervisning på vg1 svarer 69 % at de har kjennskap til begrepet utfallsrom.

Det er en forbedring på kjennskap til utfallsrom på 44 prosent mellom de to gruppene ovenfor. Resultat fra oppgave 8 viser en forbedring i prestasjonen på om lag 45 prosent. Er sammenhengen virkelig så enkel? Kanskje eller kanskje ikke, men poenget understrekes av forskere:

«Grunnleggende kunnskap i kombinatorikk er så viktig at det burde implementeres som fundamental del i sannsynlighet, og være til stede i enhver undervisningssituasjon som involverer emnet.» (Heitele, 1975. Min oversettelse).

«Mange elever vil ikke nå ønsket nivå i kombinatorikk uten spesiell undervisning i emnet.» (Fishbein og Grossman (1997) side 242 i Batanero og Sanchez, 2005. Min oversettelse).

Det er viktig at lærere prioriterer og setter fokus på en forståelse av utfallsrommet (side 228 i Tarr og Lannin, 2005). Det må altså øves på dette, eller som det heter i skolen: «Øving gjør mester»

Jeg taler varmt for at betinget sannsynlighet som er kompatibel med tidsaksen kan komme på et tidligere stadium i undervisningen. Læreplaner tar for seg betinget sannsynlighet først på videregående skole; jeg mener at der er å gjøre elevene en bjørnetjeneste ved å vente så lenge. Jeg finner noe støtte i teorien for dette punktet:

«Betinget sannsynlighet burde introduseres tidligere i læreplanene, og undervises på en intuitiv måte.» (Watson (1995) side 217 i Tarr og Lannin, 2005. Min oversettelse).

«Hvordan og på hvilket nivå betinget sannsynlighet skal komme til syne i læreplaner og undervisning kan ikke bestemmes med bakgrunn av noen få enkle undersøkelser, men må være et resultat av mange studier.» (side 217 i Tarr og Lannin, 2005. Min oversettelse).

Forskerne ovenfor sier nødvendigvis ikke på hvilket tidspunkt i skoleløpet betinget sannsynlighet skal komme inn, men jeg tolker utsagnene i lys av mine funn som understøtter en tidligere start:

6 Elevene har en intuitiv forståelse av produktsetningen

7 Prestasjoner i problemer som er kompatible med tidsaksen øker med undervisning

Det er også teori som sier at vi ikke må introdusere betinget sannsynlighet for tidlig, at alder er en faktor i forståelsen av og prestasjon med betinget sannsynlighet:

«Økende alder og undervisning er sammenfallende med økende prestasjoner i problemer som involverer betinget sannsynlighet (side 78 i Watson og Moritz, 2002).*»*

Mine undersøkelser viser at elever har en god intuitiv forståelse av asynkrone hendinger og bruk av produktsetningen, vel å merke når hendingene er kompatible med tidsaksen. Utdrag fra de enkelte læreplaner på studiespesialiserende:

- **1T: tegne valgtre og bruke addisjonssetninga og produktsetninga**
- **S1: ordnede utvalg med og uten tilbakelegging og uordnede utvalg uten tilbakelegging. Gjøre enkle sannsynlighetsberegninger knyttet til slike utvalg**
- **R1: ordnede utvalg med og uten tilbakelegging og uordnede utvalg uten tilbakelegging. Bruke dette til å utlede regler for beregning av sannsynlighet**

Vi kan jo med rette stille oss spørsmålet om det er så veldig stor forskjell på de tre punktene ovenfor. Den vanlige rekkefølgen i norsk skole er at det først undervises i snittet mellom to hendinger, før det undervises i betinget sannsynlighet. I oppgaver med kontekst er det absolutt grunn til å spørre seg hva som er lettest (side 83 i Watson og Moritz, 2002). Linda-problemet handler om snittet mellom to hendinger. Funn 3 og kjent teori bekrefter:

3 misoppfatningen conjunction fallacy, står sterkt – før og etter undervisning.

Kanskje kan vi med fordel flytte lærestoff som kan bygges rundt valgtre og produktsetningen til 1T? Betinget sannsynlighet bør komme som en naturlig del på vg1 og understrekes og styrkes på vg2, da også med Bayes' setning. Forskning viser at elever har en god og intuitiv forståelse av hendinger med og uten tilbakelegging:

"The probability changes unless you put it back in". Ungdomsskoleelev i et forsøk med kuler som ikke legges tilbake i en urne (side 216 i Tarr og Lannin, 2005).

Jeg heller til den oppfatning at det i mange sammenhenger er lettere å undervise i betinget sannsynlighet fremfor snittet mellom to hendinger. Elever på vg1 har større problemer med et venndiagram enn et valgtre. Det er en logisk tanke bak den rekkefølgen vi har i dag. Vi starter med et enkelt forsøk, før vi forflytter oss til flere suksessive forsøk. Likevel viser oppgave 2, 6a og 7, og i noen grad oppgave 11 og oppgave 12, at det er ingen særlige sterke misoppfatninger som hindrer forståelse av suksessive forsøk som er kompatible med tidsaksen.

Flytte opp eller ned? Jeg pekere på mulighetene, men mitt hovedpoeng er at plassen (tiden) som er satt av til sannsynlighetsregning i vg1 og vg2 er liten. Jeg ville hatt en større del enten på vg1 eller vg2, men helst vg1. Slik det er nå føler jeg at det er en oppdeling, som er logisk nok; tellbar mot regneregler, men som likevel på vg1 skjuler uante muligheter. Muligheter som ligger rett rundt hjørnet, men elevene må vente et år til. Et annet forslag er å vente med sannsynlighet til vg2 – og hvorfor ikke? De siste eksamensoppgavene på vg1 bærer preg av at dette blir vektlagt i mindre og mindre grad. Har det sneket seg inn en slags: «Vi må ikke gjøre oppgavene i sannsynlighetsregning for vanskelige» - tenkning i nemda? Eller er man bare gått tom for brukbare ideer? Da er mitt forslag – dropp dette i vg1, ta alt i vg2 – eller omvendt?

Hva som ikke skal flyttes ned i systemet er bayesiske problemer. Mine undersøkelser viser lav prestasjon på oppgave 13. I slike oppgaver er det viktig at informasjonen er transparent og presentert på en intuitiv måte. Med dette menes at konteksten i minst mulig grad skal forhindre forståelse, men legge til rette for forståelse. Da er det viktig at presentasjonen bruker styrken av representasjon, altså frekvenser (side 128 i Meder og Gigerenzer, 2014). Flere forskere har påpekt at prestasjonen er bedre dersom informasjonen i problemene er gitt som frekvenser i stedet for sannsynlighet (desimal eller prosent), spesielt gjelder dette problemer knyttet til Bayes' setning (side 79 i Watson og Moritz, 2002). Som en kontrast sier andre at den frekvensorienterte tilnærmingen til sannsynlighet burde fjernes, fordi dette er vanskelig uansett (Devlin, 2014). Da er det kanskje riktig at undervisningen i emnet ikke bør komme før i vg2 i R1 (realfaglig studieprogram). Kulepunkt fra læreplanen i R1:

- **gjøre rede for begrepene uavhengighet og betinget sannsynlighet, og utlede og anvende Bayes' setning på to hendelser**

Selv om frekvenser er å foretrekke, sier andre forskere at vi kan ikke slippe tak på prosenter og sannsynlighet i presentasjon av oppgaven fordi dette er viktig i mange sammenhenger (Diaz og Batanero, 2009). Min erfaring er at Bayes' setning er vanskelig og ikke bør flyttes fram, men styrkes, på vg2.

6.3.2 Utdanning

Mange lærere må tilpasse seg til en hverdag som innebærer en viss kunnskap om sannsynlighetsregning. Dessverre har flere forskere funnet at mange utdanningssituasjoner ikke gir lærere god nok innføring i emnet sannsynlighetsregning, og mange lærer manglet nødvendig utdannelse. Denne gruppen av lærere vil kanskje overføre mange av misoppfatningene til sine elever. Det er derfor viktig å evaluere lærernes kunnskap i sannsynlighetsregning og finne oppgaver som konfronterer lærere med deres egne misoppfatninger (side 346 i Batanero et al, 2014). Som eksempel vises det til misoppfatningen om at definisjonen for betinget sannsynlighet kan tolkes slik:

$$P(B|A) = \frac{Card(A \cap B)}{Card(A)}$$

Ifølge forskning er denne misoppfatningen lærerinitiert (side 252 i Batanero og Sanchez, 2005).

Sannsynlighet ble av mange lærere på videregående skole sett på som en egen side av matematikken som stort sett handlet om spill og sjanser, og ble behandlet deretter (side 124 i Batanero og Diaz, 2007). For noen år tilbake var det slik i Norge at mange matematikklærere ikke hadde formell kompetanse i sannsynlighet og statistikk. Når sannsynlighet og statistikk kom sterkere inn i læreplanene med R94, ble det på 90-tallet og utover arrangert kurs for å bøte på denne mangelen av kompetanse. Nå har det gått noen år siden 1994 og man skal ikke utelukke at lærerne faktisk har lært seg litt sannsynlighetsregning i denne perioden. I tillegg fases lærere uten formell kompetanse sakte, men sikkert ut, og erstattes av nye lærere med formell kompetanse. Prosessen går sin gang og jeg ser ingen akutte behov for å forstyrre dette bildet.

6.3.3 Undervisning

Hvordan betinget sannsynlighet presenteres er avgjørende for et godt resultat, det være seg forståelse av teori eller prestasjon i løsning av oppgaver. I sannsynlighet brukes ofte induktiv metode med eksempler som grunnlag for forståelse av teori, og nettopp derfor er sammenhengen mellom teori og oppgaver viktig.

Grunnleggende betinget sannsynlighet er tilgjengelig i tidlig alder ved bruk av enkle konkreter i enkle kontekster. Ved å trekke kuler fra en urne, kan barn i ung alder gjenkjenne med og uten tilbakelegging og gjøre adekvate beregninger (side 82 i Watson og Moritz, 2002). Forskere mener at klassiske oppgaver med kontekst, som for eksempel kuler i en urne, burde brukes som verktøy for å undersøke om elever har strukturelle kunnskaper om snitt og betinget sannsynlighet, fordi denne kunnskapen er uavhengig av kontekst. En god start kan altså være så enkelt som kuler i en urne, noe jeg ofte bruker selv (side 97-102 i Ekern med flere, 2004a). Forskere fremhever også at visuelle hjelpemidler som venndiagram, krysstabell og valgtre kan være nyttige i undervisningsøyemed (side 83 i Watson og Moritz, 2002), og at dette kan hjelpe på evnen til å forstå at utfallsrommet begrenser seg i situasjoner uten tilbakelegg (side 219 i Tarr og Lannin, 2005).

I kapittel 2 Teori, viste jeg hvordan betingelser kan ligge skjult i en tekst og hvordan en oppgave kan ha kontekstuelle hindringer. Ved å endre eller bearbeide konteksten, kan andelen som svarer riktig øke. Vi som lager oppgaver, må altså være veldig klar i formuleringen og konteksten i oppgaver generelt, og sannsynlighetsregning spesielt. Etter min erfaring lages det dessverre for mange oppgaver med kontekstuelle hindringer, slik at det blir en test på om man skjønner innholdet i teksten snarere enn det matematiske problemet (Jf. Eksamen MAT1013, V2011, oppgave 7 del 2).

Læreplanmålene nedenfor viser til eksperimenteringer og simuleringer på de to laveste nivåene, men ikke på vg2:

- Finne og diskutere sannsyn gjennom eksperimentering, simulering og berekning i daglegdagse samanhengar og spel (8. – 10. trinn)
- Formulere, eksperimentere med og drøfte uniforme og ikkje-uniforme sannsynsmodeller (vg1 – 1T)

Hva sier forskere om eksperimentering og simulering av forsøk i undervisningen? Jeg har funnet mye om dette emnet, men tar bare med et lite utvalg av hva som blir sagt, siden dette heller ikke er belyst tidligere i denne masteroppgaven. Nødvendigheten av å utføre eksperimenter, spille spill, ta i bruk simuleringer på pc for å få et tilstrekkelig empirisk grunnlag får mål på sannsynligheten av en hending er viktig. Da kan vi få ny kjennskap og hjelp til å identifisere betingede sannsynligheter.

> *«For å forbedre dyktigheten i forbindelse med sannsynlighet, må elever kjenne begrensningen til deres egen strategi i problemløsing, og lære seg å bruke verktøy utviklet av matematikere*

som hjelper i mer komplekse situasjoner» (Fishbein, 1987, side 2 i Tomlinson og Quinn, 1997).*

Simuleringer kan avsløre at ting kanskje ikke er slik som vi først trodde at det var. Et godt eksempel på dette er Monty Hall problemet der eksperimentering og simulering kan rette opp feilaktige antagelser og intuisjoner. Simuleringer oppmuntrer til tenkning og refleksjon.

Enkelte misoppfatninger består, og de står godt. Det viser kanskje at denne delen må styrkes i undervisningen? Må vi større grad utfordre elevene, eller er faren for en enda større forvirring så overhengende at vi bør la det være?

> *«Flere av misoppfatningene består selv etter undervisning, og de eksisterer i et bredt spekter. Det viser at det trengs en styrking av undervisningen i betinget sannsynlighet, og ikke minst en endring i tilnærmelsen til emnet»* (Diaz og Fuente, 2007. Min oversettelse).
>
> *«Betinget sannsynlighet bør undervises sammen med nye statistiske ideer, i presentasjon av en mengde varierende bruksområder til realistiske problemer, med aktiviteter og verdifulle presentasjoner og presentasjonsmetoder»* (Diaz og Fuente, 2007. Min oversettelse).

Det er viktig at eksperimenter og simuleringer er så varierte at de utfordrer den enkeltes heuristikk, ikke bare nører opp under en godt innarbeidet misoppfatning. Elevene må altså kjenne til heuristikker for å se sin egen begrensning. Jeg kan være enig i det, men ser også at dette må behandles varsomt.

Samarbeid mellom elevene er viktig for å undersøke og utfordre hverandres heuristikker på ulike områder innen sannsynlighet. En undersøkelse gjort av Castro (1998) sammenlignet hvor stor innvirkning to ulike typer instruksjoner påvirket elevers tenkning og prestasjon innen betinget sannsynlighet:

a) Omgivelser som fremmet og oppmuntret til tenkning og refleksjon omkring ideer om sannsynlighet.
b) Tradisjonell undervisning med en rød tråd som presentasjon av matematikken innen sannsynlighet, uten å kaste et blikk på mulige misoppfatninger.

Castro fant at misoppfatningene innen betinget sannsynlighet og uavhengighet stod sterkere og var mer motstandsdyktig i gruppen som hadde hatt tradisjonell undervisning. Disse resultatene viser viktigheten av at evaluering og instruksjon går hånd i hånd, og at instruksjon bør få ulike misoppfatninger frem i lyset, i stedet for å holde dem skjult, slik at elever får anledning til å reflektere over validiteten til deres egne intuitive tanker og heuristikker (side 229 i Tarr og Lannin, 2005).

Alle uformelle kunnskaper kan være hindre på veien til en god forståelse i følge med undervisning av sannsynlighet. For at undervisning skal ha nytte, må læreren eller underviseren stille seg inn på elevenes kanal, eller ståsted og grad av intuisjon, men også strukturere aktiviteter som legger til rette for en kognitiv konflikt og mulighet til revidering av sine forestillinger (side 152 i Konold, 1991).

1) Passer min forestilling de andre elevene sine forestillinger?
2) Er mine forestillinger konsistente med mine andre forestillinger?
3) Passer mine forestillinger med empiriske observasjoner?

Dersom elever skal tilegne seg nødvendige ferdigheter for å foreta korrekte beregninger med betinget sannsynlighet, må undervisningen fokusere på en utfordring av personlige heuristikker og kognitive heuristikker, som vist i denne oppgaven. Misoppfatningene må utfordres (side 7 i Tomlinson og Quinn, 1997). Vi trenger å ta misoppfatninger inn i vår betraktning når vi skal tilrettelegge undervisningen, i stedet for bare å konsentrere oss om algoritmer i beregning av betinget sannsynlighet og definisjoner. Dersom en elev har feilaktige ideer eller løsninger på problemer innen sannsynlighet, vil det være en ide å konfrontere eleven med eksperimenter og simuleringer. Problemene i CPR-testen og andre diagnostiske oppgaver kan være en hjelp i så måte. I starten eller innledningen av et emne bør kanskje noen diagnostiske oppgaver som leder til en kognitiv konflikt ha sin plass? Undring, eller en kognitiv konflikt oppmuntrer til spørsmål som krever svar. Min påstand er at vi som lærere i større grad må våge å utfordre elevene sine tanker innen sannsynlighet generelt og betinget sannsynlighet spesielt.

7 Appendiks og Litteratur

I dette kapitlet har jeg samlet oppgavene fra kapittel 4 og de oppgavene fra CPR som ikke var med i mine undersøkelser. En liten, eller stor, detalj er den store oversikten over «alle mulige» misoppfatninger. Ellers er det andre vedlegg og til slutt litteraturliste.

7.1 Skjema for bakgrunnsdata

I forbindelse med hver undersøkelse måtte elevene gi noe bakgrunnsinformasjon. Mitt skjema:

Bakgrunnsdata - person

1	Klasse	
2	Elevnummer	
3	Gutt / Jente	
4	Nasjonalitet	
5	Fødselsår	

Bakgrunnsdata – nivå i matematikk

6	Standpunktkarakter i matematikk 10 trinn		1T / 1P ?
7	Standpunktkarakter i matematikk vg1		
8	Eventuelt siste terminkarakter i matematikk		Nivå:

Bakgrunnsdata – tidligere undervisning i statistikk og/eller sannsynlighet

Du svarer på om du har fått undervisning på skole ifølgende emner og områder:

	Emne / Område	Ja	Nei	VI
9	Gjennomsnitt			
10	Median			
11	Typetall			
12	Variasjonsbredde			
13	Utfall og eller utfallsrom			
14	Hendelse, Hending eller Begivenhet (tre ord for samme sak)			
15	Sannsynlighet (Svar som brøk, desimal eller prosent)			
16	Telling (Hvor mange måter kan…)			
17	Har du brukt Excel eller annen programvare til å lage et diagram i statistikk?			
18	Har du gjennomført en undersøkelse i sannsynlighet eller statistikk?			
19	Kjenner du til bruken av bokstaven P når vi snakker om sannsynlighet?			

7.2 Instruksjon og Oppgavene til mine undersøkelser

Instruksjon til elevene om testen

- Det er 13 oppgaver, med deloppgaver er det totalt 18 spørsmål å besvare.
- Det er tillatt å bruke lommeregner som hjelpemiddel.
- Tillatt tid er 50 minutter.
- Du kan få vite ditt resultat, men ikke spesifikt på de enkelte oppgaver.
- Skriv på elevnummer og klasse på det siste svararket.
- Alle svar skal skrives på svararket.
- Start når du får beskjed av læreren.

Informasjonen i parentesene viser når og av hvem oppgaven er benyttet første gang. Jeg utarbeidet et eget svarark for elevene slik at overføring av data fra svararket til regnearket gikk smidig.

Oppgave 1 (Estepa, 1994)

I en medisinsk undersøkelse ble en gruppe personer intervjuet med følgende resultat:

	55 år eller yngre	Eldre enn 55 år	Total
Tidligere hjertestans	**29**	**75**	**104**
Ikke tidligere hjertestans	**401**	**275**	**676**
Total	**430**	**350**	**780**

Anta at vi velger en tilfeldig person fra denne gruppen:

e) Hvor stor er sannsynligheten for at personen har hatt hjertestans?
f) Hvor stor er sannsynligheten for at personen har hatt hjertestans og samtidig er eldre enn 55 år?
g) Dersom en person er eldre enn 55 år, hvor stor er sannsynligheten for at personen har hatt hjertestans?
h) Dersom en person har hatt hjertestans, hvor stor er sannsynligheten for at personen er eldre enn 55 år?

Oppgave 2

Det er fire lamper i en boks, to av dem er defekte. Vi plukker ut to tilfeldige lamper fra boksen, den ene etter den andre, uten å legge dem tilbake i boksen. Gitt at den første lampen er defekt; hvilket av alternativene i, ii eller iii er riktig?
Les gjennom alternativene før du svarer. Du må sette alternativene opp mot hverandre.

iv) Det er større sjanse for at den andre lampen er defekt.
v) Det er større sjanse for at den andre lampen virker.
vi) Sannsynlighetene for at den andre lampen er defekt eller at den virker, er like stor.

Oppgave 3 (analog til Tversky og Kahneman, 1982b)

En tennisspiller er i finalen i US Open. For å vinne kampen må han vinne tre av fem sett. Hvilke av følgende hendinger er mest sannsynlig eller er de alle like sannsynlige?

a) Spilleren vil vinne det første settet
b) Spilleren vil vinne det første settet, men tape kampen.
c) Begge hendinger a og b er like sannsynlige.

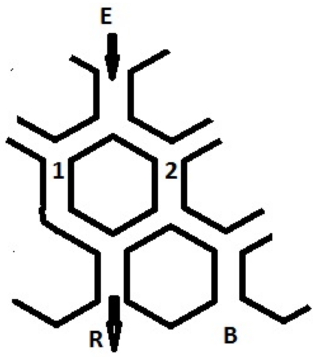
Oppgave 4 (Polatsek med flere, 1987)

A kreftundersøkelse blir gjennomført på alle innbyggerne i en by. Et positivt resultat på undersøkelsen er en indikator på kreft, mens et negativt resultat indikerer at personen er frisk. Hvilken av følgende resultater er mest sannsynlig eller de like sannsynlige?

a) Personen har kreft når testen er positiv.
b) Å ha en positiv test når personen har kreft.
c) De to hendingene a og b er like sannsynlige.

Oppgave 5 (Ojeda, 1996)

Vi slipper en ball inn gjennom åpning E på toppen i en maskin som vist på figuren. Anta at ballen forlater systemet gjennom utgang R, hvor stor er sannsynligheten for at den passerer gjennom kanal 1?



e) ½
f) 1/3
g) 2/3
h) umulig å avgjøre

Oppgave 6 (Falk, 1986)

To svarte og to hvite kuler ligger i en urne. Vi trekker ut to kuler etter hverandre fra urnen uten å legge dem tilbake. *(P betyr sannsynlighet, W betyr hvit og nummer er hvilken kule)*

a) Dersom den første kulen er hvit, hvor stor er sannsynligheten for at den andre kulen er hvit?

   $P\left(W_2 \mid W_1\right)$: i) ½ ii) 1/6 iii) 1/3 iv) ¼

b) Dersom den andre kulen er hvit, hvor stor er sannsynligheten for at den første kulen er hvit?

   $P\left(W_1 \mid W_2\right)$: i) 1/3 ii) umulig å avgjøre iii) 1/6 iv) ½

Oppgave 7

En urne inneholder en blå og to røde kuler. Vi trekker ut to kuler, en etter en, uten å legge dem tilbake. Hvilken av hendingene nedenfor er mest sannsynlig eller er de like sannsynlige?

a) Få to røde kuler
b) Den første kulen er rød og den andre kulen er blå.
c) De to hendingene a og b er like sannsynlige.

Oppgave 8

Lag utfallsrommet ferdig i de følgende tilfeldige eksperimenter:

a) Kjønnsfordelingen (Gutt/Jente) til tre barn i en familie (For eksempel GJG,…)
b) Kjønnsfordelingen (Gutt/Jente) til tre barn i en familie dersom det er to eller flere gutter.

Oppgave 9
Ved å trille to terninger (en grønn og en rød) får vi at produktet av de to tallene som terningene viser er lik 12. Hvor stor er sannsynligheten for at ingen av de to terningene er en sekser?

Oppgave 10
En person kaster en rettferdig terning og skriver ned resultatet som oddetall eller partall. Han får dette resultatet etter 15 kast: o p p o o p o o o o p p o o o
Personen kaster terningen en gang til. Hvor stor er sannsynligheten for å få et oddetall nå?

Oppgave 11
I en gruppe elever på en skole svarer 80 % av elevene at de har øvelseskjørt med foreldrene 10 ganger eller mer uten å kollidere. Av de samme elevene svarer 70 % at de bestod den siste testen i engelsk. Anta at hva en elev klarer, å bestå en test eller å øvelseskjøre uten kollisjon, er uavhengige av hverandre.
Hvor stor er da sannsynligheten for at en tilfeldig valgt elev har bestått engelsktesten og øvelseskjørt mer enn 10 ganger uten å kollidere?

Oppgave 12
Ifølge nylig forskning så lyver 91 % av folkene i en populasjon (for eksempel Bergen) og 36 % av disse lyver om viktige saker. Dersom vi plukker en tilfeldig person fra byen, hvor stor er sannsynligheten for at denne personen lyver om viktige saker?

Oppgave 13 (Totahasina, 1982)
To maskiner, M1 og M2, produserer baller. Maskin M1 produserer 40 % og maskin M2 produserer 60 % av ballene. Av ballene som blir produsert av maskin M1 er det 5 % som er defekte og av de som blir produsert av maskin M2 er det 1 % som er defekte. Vi plukker ut en tilfeldig ball og finner ut at den er defekt. Hvor stor er sannsynligheten for at ballen var produsert i maskin M1?

7.3 Utelatte oppgaver
Det var flere oppgaver fra den opprinnelige CPR-testen til Diaz med flere som var utelatt i mine undersøkelser. Jeg argumenterte for dette i kapittel 3. De utelatte oppgavene (Diaz og Batanero, 2009):

Item 2. (Tversky & Kahneman 1982a). A witness sees a crime involving a taxi in a city. The witness says that the taxi is blue. It is known from previous research that witnesses are correct 80% of the time when making such statements. The police also know that 15% of the taxis in the city are blue, the other 85% being green. What is the probability that a blue taxi was involved in the crime?

a. $\frac{80}{100}$ b. $\frac{15}{100}$ c. $\frac{15}{100} \times \frac{80}{100}$ d. $\frac{15 \times 80}{85 \times 20 + 15 \times 80}$

Item 3. (Sánchez 1996). A standard deck of playing cards has 52 cards. There are four suits (clubs, diamonds, hearts, and spades), each of which has thirteen cards (2, ... , 9, 10, *Jack*, *Queen*, *King*, *Ace*). We pick a card up at random. Let A be the event "*getting diamonds*" and B the event "*getting a Queen*". Are events A and B independent?

a. A and B are not independent, since there is the *Queen* of *diamonds*.
b. A and B are only then independent when we first get a card to see if it is a diamond, return the card to the pack and then get a second card to see if it is a *Queen*.
c. A and B are independent, since $P(Queen \text{ of } diamonds) = P(Queen) \times P(diamonds)$.
d. The events A and B are not independent, since $P(Queen \mid diamonds) \neq P(Queen)$.

Item 5. (Eddy 1982). 10.3 % of women in a given city have a positive mammogram. The probability that a woman in this city has both positive mammogram and breast cancer is 0.8%. A mammogram given to a woman taken at random in this population was positive. What is the probability that she actually has breast cancer?

a. $\frac{0.8}{10.3} = 0.07767$, $7.77\ \%$ b. $10.3 \times 0.8 = 8.24\ [\%]$, $8.24\ \%$ c. $0.8\ \%$

Item 11. Explain in your own words what a simple and a conditional probability are and provide an example.

Item 14. 60% of the population in a city are men, 40% women. 50% of the men and 25% of the women smoke. We select a person from the city at random; what is the probability that this person is a smoker?

Avsnitt 7.2 og 7.3 gir til sammen den komplette CPR-testen til Diaz med flere (Diaz og Batanero, 2009).

7.4 Liste over misoppfatninger

Etter å ha fått mer og mer kunnskap om emnet misoppfatning, ble jeg også mer interessert i emnet. Jeg tok derfor med denne listen fra Wikipedia – og jammen dukker det opp en IKEA her også.

Ambiguity effect, Anchoring or focalism, Attentional bias, Automation bias, Availability heuristic, Availability cascade, Backfire effect, Bandwagon effect, Base rate fallacy or Base rate neglect, Belief bias, Bias blind spot, Cheerleader effect, Choice-supportive bias, Clustering illusion, Confirmation bias, Congruence bias, Conjunction fallacy, Regressive bias, Conservatism (Bayesian), Contrast effect, Curse of knowledge, Decoy effect, Denomination effect, Disposition effect, Distinction bias, Dunning-Kruger effect, Duration neglect, Empathy gap, Endowment effect, Essentialism, Exaggerated expectation, Experimenter's or expectation bias, Focusing effect, Forer effect or Barnum effect, Framing effect, Frequency illusion, Functional fixedness, Gambler's fallacy, Hard–easy effect, Hindsight bias, Hot-hand fallacy, Hyperbolic discounting, Identifiable victim effect, IKEA effect, Illusion of control, Illusion of validity, Illusory correlation, Impact bias, Information bias, Insensitivity to sample size, Irrational escalation, Less-is-better effect, Loss aversion, Mere exposure effect,

Money illusion, Moral credential effect, Negativity effect, Negativity bias, Neglect of probability, Normalcy bias, Not invented here, Observer-expectancy effect, Omission bias, Optimism bias, Ostrich effect, Outcome bias, Overconfidence effect, Pareidolia, Parkinson's Law of Triviality, Pessimism bias, Planning fallacy, Post-purchase rationalization, Pro-innovation bias, Pseudocertainty effect, Reactance, Reactive devaluation, Recency illusion, Restraint bias, Rhyme as reason effect, Risk compensation / Peltzman effect, Selective perception, Semmelweis reflex, Social comparison bias, Social desirability bias, Status quo bias, Stereotyping, Subadditivity effect, Subjective validation, Survivorship bias, Time-saving bias, Unit bias, Weber–Fechner law, Well-travelled road effect, Zero-risk bias, Zero-sum heuristic
Kilde: https://en.wikipedia.org/wiki/List_of_cognitive_biases

Dersom du ønsker å vite hva «IKEA effect» er for noe – så gå til kilden.

7.5 Læreplanene for S1 og R1

Fagene S1 og R1 er omtalt i masteroppgaven, men de tilhørende læreplanene for emnet sannsynlighetsregning er ikke tatt med tidligere. Her kommer de:

Læreplan S1

- regne med binomialkoeffisienter og bygge opp Pascals talltrekant
- gjøre rede for ordnede utvalg med og uten tilbakelegging og uordnede utvalg uten tilbakelegging, og gjøre enkle sannsynlighetsberegninger knyttet til slike utvalg
- lage binomiske og hypergeometriske sannsynlighetsmodeller ut fra praktiske situasjoner, og regne med sannsynligheter for slike modeller

Læreplan R1

- gjøre rede for begrepene uavhengighet og betinget sannsynlighet, og utlede og anvende Bayes' setning på to hendelser
- drøfte kombinatoriske problemer knyttet til ordnede utvalg med og uten tilbakelegging og uordnede utvalg uten tilbakelegging, og bruke dette til å utlede regler for beregning av sannsynlighet

7.6 Transkripsjon

Jeg har gjennomført fire intervjuer på om lag 30 – 40 minutter hver. Intervjuene eksisterer som lyd-filer og er i tillegg transkribert. Lydfilene slettes straks oppgaven er levert. Transkripsjonen er ikke med som vedlegg i dette kapittelet. Utklippene nedenfor viser til lydfilene og transkripsjonen.

Intervju 1 Pilot kand 1 3 5 10032015 Mast...	Digitallyd	97 215 kB
Intervju 2 Und 1 1G 5 13 17 19032015 Mas...	Digitallyd	81 725 kB
Intervju 3 Und 2 1G 8 10 16 19032015 Mas...	Digitallyd	82 563 kB
Intervju 4 Enkel elev S1 220415 Master Ui...	Digitallyd	71 383 kB

Utklipp: Lydfiler fra fire intervju.

Transkripsjon Intervju 4 Enkel Elev S1 220...	30.05.2015 09:54	Microsoft Word-d...	38 kB
Transkripsjon Intervju 3 Und 2 190315 ele...	30.05.2015 09:50	Microsoft Word-d...	38 kB
Transkripsjon Intervju 2 Und 1 190315 ele...	30.05.2015 09:45	Microsoft Word-d...	38 kB
Transkripsjon Intervju 1 Pilot 100315 elev...	30.05.2015 09:38	Microsoft Word-d...	37 kB

Utklipp: Transkripsjon fra fire intervju.

7.7 Litteraturliste

Jeg har jobbet for å få med sidetall i alle referanser, men har ikke lyktes helt. Noen artikler er faktisk uten sidetall og noen artikler er bare på et par sider. I disse tilfellene har jeg droppet henvisning til sidetall. I tillegg har jeg droppet sidetall der disse er nummerert med romertallsystemet (jamfør Devlin, 2014 og Chernoff og Sriraman, 2014). I et enkelt tilfelle har jeg telt sidetallet fra start (jamfør Tomlinson og Quinn, 1997, som starter på side 2 og slutter på side 7). I noen tilfeller har jeg ikke funnet tilbake til sidene jeg referer til, men jeg håper at referanser uten sidetall er i absolutt mindretall. Jeg har ikke referert til sidetall for Batanero og Diaz (2009) og Diaz og Fuente (2007), fordi artiklene er så sentralt i denne masteroppgaven at alt bør leses.

Falk, R. (1986) 'Conditional Probabilities: Insight and Difficulties', ICOTS 2

Diaz, C. og Batanero, C. (2009) 'University students' knowledge and biases in conditional probability reasoning', International electronic journal of mathematics education, Vol. 4, no. 3

Watson, J. M (2006) 'Issues for statistical literacy in the middle school', ICOTS-7, 2006

Madsen, R. W. (1995) 'Secondary students' Concepts of Probability',Teaching Statistics. An International Journal for Teachers, Vol. 33, No. 3, pp. 82-112

Batanero, C. og Sanchez, E. (2005) 'What is the nature of high school students' conceptions and misconceptions about probability?', Exploring probability in school: Challenges for teaching and learning, chapter 10, pp. 241 – 266

Watson, J. M. og Moritz, J. B. (2002) 'School students' reasoning about conjunction and conditional events', International Journal of Mathematical Education in Science and Technology, Vol. 33, No. 1, pp. 59 – 84

Batanero, C. og Diaz, C. (2007) 'The meaning and Understanding of Mathematics. The Case of probability', Philosophical dimensions in mathematics education. Mathematics education library 42, pp. 107 – 127

Konold, C. (1991) 'Understanding students' beliefs about probability', Radical constructivism in Mathematics Edication, pp. 139 – 156

Tomlinson, S. og Quinn, R (1997) 'Understanding conditional Probability', Teaching Statistics. An International Journal for Teachers, Vol. 19, No. 1, pp. 2-7

Tarr, J. E. og Lannin, J. K. (2005) 'How can teachers build notations of conditional probability and independence?, Exploring probability in school: Challenges for teaching and learning, chapter 9, pp. 215 – 228

Falk, R. og Kendig, K (2012) 'A tale of two probabilities', Teaching statistics trust, 35, 1, 49 – 52.

Polaki, M. V. (2005) 'Dealing with compound events', Exploring probability in school: Challenges for teaching and learning, chapter 8, 191 – 214

Devlin, K. (2014) 'The Most Common Misconseption About Probability', Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Chernoff, E. J. og Sriraman, B. (2014) ‘Introduction to Probabilistic Thinking: Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Chernoff, E. J. og Russell, G. L. (2014) ‘Preface to Perspective I: Mathematics and Philosophy’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Borovcnik, M. og Kapadia, R (2014) ‘A Historical and Philosophical Perspective on Probability’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business.

Pfannkuch, M. og Ziedins, I. (2014) ‘A Modelling Perspective on Probability’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business.

Goldin, G. A. (1993) ‘Observing Mathematical Problem Solving: Perspectives on Structured, Task-based Interviews’.

Johnson, R. B og Onwuegbuzie, A. J. (2004) ‘Mixed Methods Research: A Reasearch Paradigm Whose Time Has Come’, Educational Researcher, Vol. 33, No. 7, pp. 14-26.

Gellert, U., Hernandez, R. B. og Chapman, O. (2013) ‘Research Methods in Mathematics Teacher Education’, Third International Handbook of Mathematics Education. Chapter 11, Clements,. M. A. med flere (redaktør), Springer International Handbooks of Education 27

Skovlund, E. (2015) ‘Et uunnværlig verktøy’, Statistikk og medisin. Tidsskrift for Den norske legeforening, 16-2015, p. 1424

Ekern, T., Guldahl, Ø., Holst, E. (2004a) ‘Paralleller. 2mx’, Kapittel 6, NKI Forlaget

Ekern, T., Guldahl, Ø., Holst, E. (2004b) ‘Paralleller. 1mxy’, Kapittel 3, NKI Forlaget

Martyn, C. (2014) ‘Risky business: doctors’ understanding of statistics’, BMJ 2014; 349:g5619

Lecoutre, M.-P., (1992) ‘Cognitive Models and Problem Spaces in “Purely Random” Situations’, Educational Studies in Mathematics, Vol. 23, No. 6, pp. 557 – 568

Fishbein, E., Nello, M. S., Marino, M. S., (1991) ‘Factors Affecting Probabilistic Judgements in Children and Adolescents’, Educational Studies in Mathematics, Vol. 22, No. 6, pp. 523 – 549

Stanovich, K. E., West, R. F., (2000) ‘Individual Differences in Reasoning: Implications for the Rationality Debate?’, Behavioral and Brain Sciences, Vol. 23, pp 645 – 726

Shleifer, A., (2012) ‘Psychologists at the gate: review of Daniel Kahneman’s Thinking, Fast and Slow’, Journal of Economic Literature 50(4), pp. 1080 – 1091

Dooren, W. V., (2014) ‘Analyses from a Psychological Perspectives’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Meder, B., Gigerenzer, G., (2014) ‘Statistical Thinking: No one left behind’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Batanero, C., Arteaga, P., Serrano, L., Ruiz, B., (2014) ‘Prospective Primary School Teachers’ Perception of Randomness’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Savard, A., (2014) ‘Developing Probabilistic Thinking; What About People’s Conceptions?’ Perception of Randomness’, Probabilistic Thinking: Presenting Plural Perspectives, Advances in Mathematics Education, Springer Science+Business

Tversky, A., Kahneman, D., (1974) ‘Judgment under Uncertainty: Heuristics and Biases’, Science. New Series, Vol. 185, No. 4157, pp. 1124 – 1131, (gjengitt I Probabilistic Reasoning, chapter 3)

Tversky, A., Kahneman, D., (1983) 'Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement', Psychological Review, Volume 90, Number 40, October 1983

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., Lipson, A., (1993) 'Inconsistent in Students' Reasoning about Probability', Journal for Research in Mathematics Education, Vol. 24, No. 5, pp. 392 – 414

Diaz, C. de la Fuente, I., (2007) 'Assessing Students' Difficulties with Conditional Probability and Bayesian Reasoning', International electronic journal of mathematics education, Vol 2, no 3, October 2007

Bryant, P., Nunes, T., (2012) 'Children's understanding of probability', A literature review (full report), Nuffield Foundation.

Brekke, G., (2002) 'Introduksjon til diagnostisk undervisning I matematikk: Kartlegging av matematikkforståelse', Læringssenteret (LS)

Store norske leksikon (lenke fra kapittel 1.5.1)

Creswell, J. W. (2012) 'Educational Research. Planning, Conducting, and Evaluating Quantitative and Qualitative research', Pearson

Goldin, G. A. (2000) 'A Scientific Perspective on Structured, Tasked-Based Interviews in Mathematical Education Research', Handbook for Research Design in Mathematics and Science Education; 517 – 545, Editor: Kelly, A. E. og Lesh, R. A., (2000).

Thoresen, S. B., (2015) 'Norske forskere er for dårlige I statistikk', Aftenposten, 5/3-15; p. 35

Devlin, K., (2008) 'The Unfinished Game: Pascal, Fermat, and the Seventeenth Century Letter that Made the World Modern', Basic Books.

Pripp, A. H. (2015) 'Hvorfor p-verdien er signifikant. Statistikk og medisin', Tidsskrift for Den norske legeforening, 16-2015; 1462-1464.

Brakedal, B. (2015) 'En kliniker og en bayesianer'. Tidsskrift for den norske legeforening, 16-2015; 1468-1470

Lillestøl, J. (1991) 'Sannsynlighetsregning og statistikk. Med anvendelser', Bedriftsøkonomens Forlag. Kapittel 4, pp. 87-89