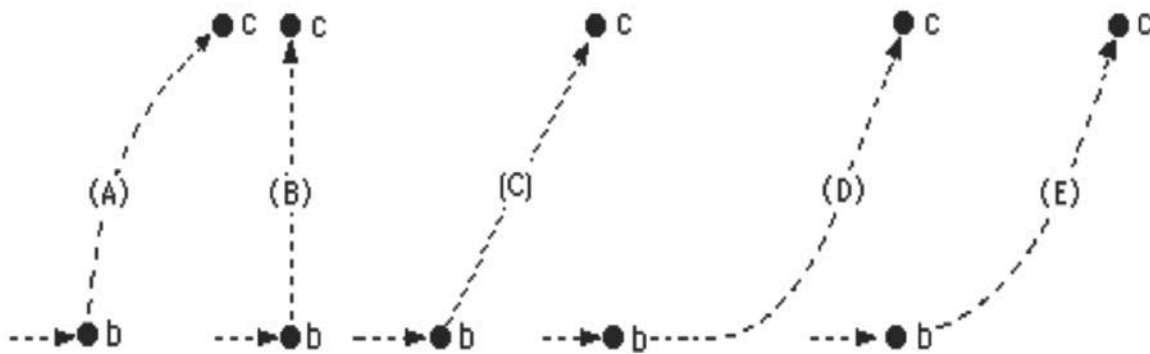




## Køn, evaluering og The Force Concept Inventory



Sofie Birch Jensen  
Kandidatspeciale

April 2013

IND's studenterserie nr. 30

## IND's studenterserie

1. Ellen Berg Jensen: 15-åriges viden om klimaforskelle (2007)
2. Martin Sonnenborg: The Didactic Potential of CAS (2007)
3. Karina Søgaard og Sarah Kyhn Buskbjerg: Galoisteori i Gymnasiet (2007)
4. Ana Hesselbart: Mathematical reasoning and semiosis (2007)
5. Julian Tosev: Forskningslignende situationer (2007)
6. Niels Nørskov Laursen: En Covarians-tilgang til Variabelssammenhænge i gymnasiet (2007)
7. Katja Vinding Petersen: Lyd og Liv (2007)
8. Jesper Bruun: Krop og computer i fysikundervisning (2008)
9. Jakob Svendsen: Matematiklærerens forberedelse (2009)
10. Britta Hansen: Didaktik på tværs af matematik og historie (2009)
11. Nadja Ussingkær: En didaktisk undersøgelse af brudte lineære funktioner i rammerne af tysk fritidsmatematik (2009)
12. Thomas Thrane Design og test af RSC-forløb om vektorfunktioner og bevægelse
13. Flemming Munch Hansen: Samspil omkring differentialregningens elementer i gymnasiets matematik og fysik (2009)
14. Hasan Ademovski og Hatice Ademovski: Proportionalitet på mellemtrinnet - Design af didaktiske situationer baseret på stofdidaktisk analyse (2009)
15. Philipp Lorenzen: Hvem er de nye studenter? Baggrund, interesse & uddannelsesstrategi (2010)
16. Signe Ougaard: Logiske strukturer i matematisk analyse på gymnasieniveau. Et forløb om kvantorer og  $\epsilon\delta$ -definition af grænseværdi (2010)
17. Jesper Winther Sørensen: Abstrakt algebra i gymnasiet - design, udførelse og analyse af undervisning i gruppeteori (2010)
18. Sofie Stoustrup: En analyse af differentiallyigninger på A-niveau i STX ud fra den antropologiske didaktiske teori (2010)
19. Jan Henrik Egballe Heinze: Eksponentialfunktioner i STX (2010)
20. Mette Beier Jensen: Virtuelgalathea3.dk i biologiundervisningen i gymnasiet (2010)
21. Servet Dönmez: Tosprogede elever og matematik i gymnasiet (2011)
22. Therese Røndum Frederiksen: Designing and implementing an engaging learning experience about the electric sense of sharks for the visitors at Danmarks Akvarium (2011)
23. Sarah Elisabeth Klein: Using touch-tables and inquiry methods to attract and engage visitors (2011)
24. Line Kabell Nissen: Matematik for Sjøv – Studie- og forskningsforløb i en eksperimentel kontekst (2011)
25. Jonathan Barrett: Trigonometriske Funktioner i en Gymnasial Kontekst – en epistemologisk referencemodel (2011)
26. Rune Skalborg Hansen: Et studie i læringsopfattelse og -udbytte - om fysik C kursister på Københavns VUC (2011)
27. Astrid Camilus: Valideringssituationer i undervisningsforløb om differentiallyigninger og radioaktivitet (2012)
28. Niven Adel Atie: Didaktiske situationer for fuldstændiggørelse af kvadratet i andengrads ligningen (2013)
29. Morten C. B. Persson: Kvantekemi i gymnasiet - Tilrettelæggelse, udførelse og evaluering af et undervisningsforløb (2013)
30. **Sofie Birch Jensen: Køn, evaluering og The Force Concept Inventory (2013)**

## Abstract

Denne specialrapport omhandler FCI-testen, en hyppigt anvendt test af begrebsforståelse i Newtonsk mekanik. I undersøgelsens første del analyseres FCI-data fra Niels Bohr Institutet kvantitativt med henblik på at undersøge, om en i udlandet observeret kønsforskel i besvarelserne genfindes i en dansk kontekst. I undersøgelsens anden del er 12 studerendes forståelse af udvalgte FCI-spørgsmål testet i et sociokulturelt orienteret evalueringsformat med det formål at undersøge, hvilken betydning evalueringsformatet har for studerendes besvarelser. Den til det formål indsamlede empiri giver desuden anledning til en læringsteoretisk analyse af, hvordan studerende tilegner sig forståelse af Newtonsk mekanik på en måde der betyder, at deres forståelse er kontekstafhængig.

*IND's studenterserie består af kandidatspecialer og bachelorprojekter skrevet ved eller i tilknytning til Institut for Naturfagenes Didaktik. Disse drejer sig ofte om uddannelsesfaglige problemstillinger, der har interesse også uden for universitetets mure. De publiceres derfor i elektronisk form, naturligvis under forudsætning af samtykke fra forfatterne. Det er tale om studenterarbejder, og ikke endelige forskningspublikationer. Se hele serien på: [www.ind.ku.dk/publikationer/studenterserien/](http://www.ind.ku.dk/publikationer/studenterserien/)*

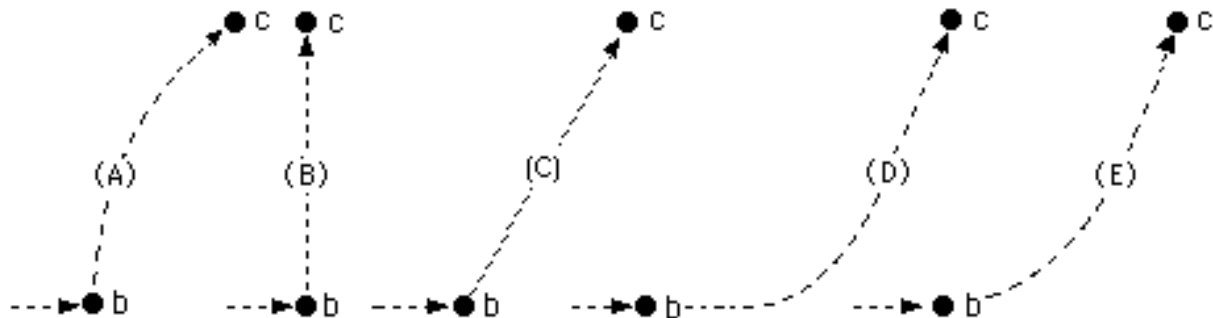


# Køn, evaluering og The Force Concept Inventory

Kandidatspeciale

Sofie Birch Jensen

Januar 2013



## Vejledere

Lene Møller Madsen, Institut for Naturfagernes Didaktik, Københavns Universitet  
Ian Bearden, Niels Bohr Institutet, Københavns Universitet

## Abstract

The results of The Force Concept Inventory, a often used test of the conceptual understanding of Newtonian mechanics, show a persistent gender gap. This is surprising since the same tendency is not found in related exam results. In this thesis this gender gap is investigated in a Danish setting.

In the first part of the study, FCI data from the Niels Bohr Institute (Department of Physics at The University of Copenhagen) from the years 2008 to 2011 is analyzed using quantitative methods. It is concluded that the average male score is significantly higher than the average female score.

The same data set is analyzed in order to investigate if the gender gap can be explained through gender dependent tendencies to leave questions unanswered when in doubt. It is concluded that such a difference exists in the data (more females than males leave questions unanswered) but the number of questions left unanswered is too small to explain the gender gap.

In the second part of the study, 12 first year students from The Niels Bohr Institute were chosen to participate in a socio-culturally orientated assessment of their knowledge regarding 6 of the FCI questions. It is seen that the change of the assessment form influences the answers given by the students. It is thus concluded that the perspective on the students' knowledge regarding Newtonian mechanics offered by the FCI is dependent on the choice of assessment form. It is furthermore concluded that the qualitative analysis doesn't suggest that this change in answers is gender dependent and thus able to explain the gender gap in the FCI in its original form. However the data set is not large enough to statistically reject that the gender gap would change if a larger population was to be tested using the socio-culturally orientated assessment form.

The new answers to one of the FCI questions regarding parabolic orbits are analyzed thoroughly. It is concluded that the students' understanding of the physics necessary to answer the question is dependent of the context of the question and that the understanding of the FCI results should reflect this context dependency.

**Keywords:** Force Concept Inventory, physics education research, socio-cultural assessment, multiple choice testing and item omission, gender

## Sammendrag

Der eksisterer en velunderbygget kønsforskel i resultaterne af The Force Concept Inventory-testen, en hyppigt anvendt test af begrebsforståelse i Newtonsk mekanik. Denne forskel er overraskende, da den ikke genfindes i eksamensresultater inden for samme fagområde. I dette speciale undersøges denne kønsforskel i en dansk kontekst.

I undersøgelsens første del er FCI-data fra Niels Bohr Institutet på Københavns Universitet fra årene 2008 til 2011 analyseret kvantitativt. Det konkluderes, at mænds gennemsnitlige procentuelle score er højere end den gennemsnitlige score for kvinderne.

De samme FCI-data er analyseret med henblik på at undersøge, om kønsforskellen kan forklares med en hyppigere tendens til at udelade svar i tilfælde af usikkerhed hos kvinderne end hos mændene. Det viser sig, at der er signifikant forskel på kønnenes tendens til svarudeladelser (der er flere kvinder end mænd der udelader svar), men at der ikke udelades nok svar til, at det kan være en fyldestgørende forklaring af kønsforskellen.

I undersøgelsens anden del blev 12 studerende fra årgang 2012 på Niels Bohr Institutet udvalgt til at deltage i en sociokulturelt orienteret evaluering af deres forståelse af 6 FCI-spørgsmål. Det viser sig, at ændringen af evalueringsformatet fra det oprindelige multiple choice-format har en indflydelse på de studerendes besvarelser, der i mange tilfælde ændres og det konkluderes, at det indblik, FCI-testen giver i studerendes forståelse af Newtonsk mekanik hænger sammen med valget af evalueringsformat. Det konkluderes yderligere, at der ikke ud fra den kvalitative analyse ses tegn på, at ændringen er kønsspecifik og dermed kan forklare kønsforskellen i FCI-scoring i det oprindelige evalueringsformat. Omvendt er datamaterialet ikke stort nok til statistisk at afvise, at kønsforskellen ville ændres hvis en hel årgang blev testet i det sociokulturelt orienterede evalueringsformat.

Besvarelserne af et enkelt af FCI-spørgsmålene omhandlende parabelbaner analyseres særlig grundigt. Det konkluderes, at de studerendes forståelse af den bagvedliggende fysik er kontekstafhængig og at den læringsteoretiske forståelse af FCI-testens resultater bør søge at afspejle dette.

**Nøgleord:** Force Concept Inventory, køn, fysikdidaktik, sociokulturel evaluering, svarstrategier i multiple choice-tests

## Forord

Denne afhandling udgør det afsluttende kandidatspeciale i uddannelsen som cand. scient. i fysik og matematik fra Det Naturvidenskabelige Fakultet på Københavns Universitet og er udarbejdet ved Institut for Naturfagenes Didaktik i efteråret 2012.

Jeg vil gerne sige tak til Institut for Naturfagenes Didaktik for at stille en specialeplads til rådighed og for åbenhed og imødekommenhed. Desuden tak til min kontormakker Rie for gode råd og kaffepauser undervejs i skrivningen.

Herudover vil jeg gerne sige tak til mine venner og Elersianerne for opbakning undervejs i skriveprocessen. En særlig tak til Anne-Katrine for hjælp med korrekturlæsning, udprintning og indbinding.

Slutteligt vil jeg gerne rette en stor tak til mine vejledere Lene Møller Madsen, Institut for Naturfagenes Didaktik og Ian Bearden, Niels Bohr Institutet for inspirerende vejledning og interessante diskussioner.

København, 10. januar 2013

Sofie Birch Jensen

# Indholdsfortegnelse

Abstract.....	2
Sammendrag .....	3
Forord.....	4
<b>1 Indledning.....</b>	<b>8</b>
<b>2 Læringsteori.....</b>	<b>10</b>
2.1 Hvad er læringsteori?.....	10
2.2 Behaviorisme .....	10
2.3 Konstruktivisme .....	11
2.3.1 Piaget og mentalkonstruktivisme.....	11
2.3.2 Vygotsky og socialkonstruktivisme.....	12
2.4 Opsummering og sammenligning af Piaget og Vygotsky.....	14
<b>3 Evaluering.....</b>	<b>15</b>
3.1 Begrundelser for evaluering.....	15
3.2 Formativ og summativ evaluering.....	16
3.3 Evalueringens validitet og reliabilitet.....	16
3.4 Evalueringsparadigmer og læringsteori.....	17
<b>4 Køn .....</b>	<b>19</b>
<b>5 Force Concept Inventory.....</b>	<b>20</b>
5.1 Hvad er FCI? .....	20
5.2 FCI-gain .....	21
5.3 FCI's anvendelsesmuligheder .....	21
5.4 FCI's validitet og reliabilitet.....	21
5.5 Tidligere kritik af FCI .....	22
5.6 Lærings- og evalueringsteoretisk analyse af FCI .....	23
5.6.1 Læringsteoretisk analyse.....	23
5.6.2 Evalueringsteoretisk analyse .....	25
<b>6 Kønsforskellen i FCI .....</b>	<b>26</b>
6.1 Kønsforskellen i udlandet .....	26
6.2 Kønsforskellen i FCI – tidligere forklaringsmodeller.....	27
<b>7 Analyse af FCI-data fra NBI.....</b>	<b>30</b>
7.1 Kvantitativ metode.....	30
7.2 Analyse af scorerer fordelt på køn.....	31
7.3 Analyse af kønsforskellen fordelt på de enkelte spørgsmål.....	34
7.4 Analyse af udeladte svar .....	36
7.5 Konklusion.....	38
<b>8 Gentestning af FCI-spørgsmål – empiriindsamling og metode.....</b>	<b>39</b>
8.1 Validering af PISA-projekt.....	39
8.2 Case studies .....	40
8.2.1 Case studies og køn.....	42
8.3 Udvælgelsen af cases .....	42
8.4 Udvælgelse af spørgsmål.....	44
8.5 Sociokulturelle aspekter .....	45
8.6 Interviewenes struktur .....	46

8.7	Forskningsinterview eller evalueringssamtale?	47
8.8	Dataindsamling	47
8.9	Metodiske vanskeligheder	48
<b>9</b>	<b>Resultater</b>	<b>50</b>
9.1	Sammenligning af svar	50
9.2	Nuancering af scorer	53
9.3	Konklusion	55
<b>10</b>	<b>Evalueringsteoretisk analyse af den udførte empiri</b>	<b>56</b>
10.1	Anvendelse af de materielle artefakter	56
10.2	Anvendelse af sproget som artefakt	57
10.3	Viden som socialt situeret	59
10.4	Viden som kontekstuel situeret	59
10.5	Tids- og læsefærdighedsproblematikker	60
10.6	Grupper af sammenhængende spørgsmål	61
10.7	Stilladsering	61
10.8	Gentestningsproblematikken	63
10.9	Konklusion	64
<b>11</b>	<b>Læringsteoretisk analyse af den udførte empiri</b>	<b>65</b>
11.0.1	Spørgsmål 21 besvaret korrekt og fyldestgørende	66
11.0.2	Spørgsmål 21 besvaret korrekt, men ikke fyldestgørende	66
11.0.3	Spørgsmål 12 besvaret korrekt og fyldestgørende	67
11.0.4	Spørgsmål 12 besvaret korrekt, men ikke fyldestgørende	68
11.1	Hvordan bygges bro mellem de to spørgsmål?	68
11.2	Hvordan fortolkes ovenstående læringsteoretisk?	70
11.3	Konklusion	73
<b>12</b>	<b>Diskussion</b>	<b>74</b>
12.1	Evalueringsteoretisk diskussion	74
12.2	Læringsteoretisk diskussion	75
<b>13</b>	<b>Perspektivering</b>	<b>77</b>
13.1	Kønssforskellen i FCI	77
13.2	Videre anvendelser af den indsamlede empiri	78
<b>14</b>	<b>Konklusion</b>	<b>80</b>
<b>15</b>	<b>Litteraturliste</b>	<b>82</b>
<b>Appendix A</b>	<b>Force Concept Inventory-testen i dansk oversættelse</b>	<b>86</b>
<b>Appendix B</b>	<b>Tabeller over FCI-spørgsmål</b>	<b>96</b>
<b>Appendix C</b>	<b>Fysikfaglig analyse af udvalgte spørgsmål</b>	<b>99</b>
	Spørgsmål 13 og 30	99
	Spørgsmål 21, 22, 23 og 24	99
	<b>Udarbejdelsen af samtaleskemaet</b>	<b>99</b>
	Spørgsmål 13	99
	Spørgsmål 30	100
	Spørgsmål 21-24	100



<b>Appendix D.....</b>	<b>102</b>
<b>    Samtaleskema.....</b>	<b>102</b>
<b>    Pointgivningsskema .....</b>	<b>105</b>

## 1 Indledning

Et speciale, der er frugten af et halvt års arbejde, afspejler på mange måder en lang proces. Under udarbejdelsen af specialet og i forbindelse med dataindsamlingen er jeg blevet klogere på det problemfelt jeg har undersøgt – hvilket også betyder, at problemfeltet har ændret sig i takt hermed.

Mit oprindelige interesseområde var en undersøgelse af overgangen fra ungdomsuddannelse til et kønsskævt studium, som fysikstudiet på Københavns Universitet er. Jeg var interesseret i, hvordan de studerende håndterede denne overgang og hvordan deres identitetskonstruktion som fysikstuderende blev influeret af deres køn. Indgangsvinklen til at behandle dette brede problemfelt blev Force Concept Inventory-testen – en test, der gives til de studerende ved studiestart og som har til formål at evaluere deres konceptuelle forståelse af Newtonsk mekanik. Det viser sig nemlig overraskende nok, at der på universiteter i udlandet er en markant forskel på henholdsvis kvindelige og mandlige studerendes FCI-score, hvor mandlige studerende gennemsnitligt scorer højere end kvindelige. Jeg startede derfor arbejdet med mit speciale med det formål at undersøge denne kønsforskel i en dansk kontekst – eksisterer kønsforskellen også her og hvad kan den skyldes? For at forstå årsagerne til kønsforskellen blev der opstillet en arbejdshypotese om, at kønsforskellen kunne hænge sammen med FCI-testens multiple choice-format og at det derfor ville være interessant at forsøge at gennemføre en tilsvarende test i et ændret evalueringsformat for at se, om dette ville påvirke kønsforskellen. Efterhånden som jeg fik indsamlet og bearbejdet data til at efterprøve denne hypotese viste det sig dog, at ændringen af evalueringsformen medførte resultater, der var særdeles interessante i deres egen ret og ikke kun som en mulig forklaringsmodel på en observeret kønsforskel. Således er det kønsfokus i relation til FCI-testen, som specialet havde i starten af arbejdsprocessen, langsomt blevet erstattet af et fokus på evaluering. Man vil derfor opleve to gennemgående temaer, køn og evaluering, der ligger som røde tråde gennem hele specialet, men hvor kønsspørgsmålet efterhånden træder i baggrunden for at erstattes af et fokus på evaluering.

De to overordnede forskningsspørgsmål, som jeg forsøger at besvare med udarbejdelsen af dette speciale kan dermed opsummeres i følgende formål:

- At undersøge, hvorvidt den i udlandet observerede kønsforskel genfindes på Niels Bohr Institutet og at undersøge hvilken rolle de studerendes blanke svar har i forhold til FCI-resultaterne og at undersøge, om der ses en kønsforskel her der kan være med til at forklare den observerede kønsforskel i FCI.
- At undersøge, hvad der sker med resultaterne af FCI-testen hvis evalueringsformen ændres fra en traditionel multiple choice-test til en sociokulturelt orienteret evaluering og at forstå, hvad denne resultatændring skyldes.

I processen med at analysere den indsamlede empiri i form af besvarelser af udvalgte FCI-spørgsmål i et nyt evalueringsformat blev jeg bevidst om nogle perspektiver i empirien, som var for væsentlige til at undlade at beskæftige sig med selvom de ikke direkte relaterer til nogen af de opstillede forskningsspørgsmål. Det viste sig nemlig, at de interviewede studerendes forståelse af de fysiske koncepter i et af spørgsmålene ændrede sig voldsomt når disse koncepter blev udfoldet i et andet spørgsmål med en anden kontekst. Således bliver en

væsentlig del af analysen i, og dermed også formålet med, dette speciale at søge at forstå denne forskel i forståelse ud fra et læringsteoretisk perspektiv.

Specialet er opbygget således, at der først gives en introduktion til det teoretiske fundament, som specialet bygger på. Der gives således en kort introduktion til læringsteori, der dels er relevant i forhold til den senere analyse, men som også fungerer som et grundlag for at beskrive, hvordan forskellige læringsteoretiske udgangspunkter influerer på opfattelsen af evaluering. Dette munder ud i en definition af to evalueringsparadigmer. Herudover gives der en kort introduktion til, hvordan køn opfattes i andre videnskabelige arbejder der beskæftiger sig med fysik og køn og specialet sættes i relation hertil.

Efterfølgende beskrives FCI-testen og det opstillede teoretiske fundament bruges til en analyse af, hvilket lærings- og evalueringsteoretisk perspektiv FCI-testen sædvanligvis betragtes i. Ligeledes diskuteres det, hvordan kønsforskellen i FCI-testen tidligere er beskrevet i den fysikdidaktiske litteratur.

Efter dette fundament er etableret kan den egentlige analyse begynde. Denne falder i to dele, relateret til hver sit forskningsspørgsmål og med grundlæggende forskellige metoder. Således gennemgås først en analyse relateret til specialets første forskningsspørgsmål, som er foretaget ved hjælp af kvantitativ metode. Resultaterne af denne analyse bidrager med perspektiver på, hvordan specialets andet forskningsspørgsmål skal gribes an. Det metodiske grundlag for og udførelsen af den gennemførte empiri præsenteres herefter. Denne empiri analyseres herefter på to måder: først ud fra et evalueringsteoretisk perspektiv og herefter ud fra et læringsteoretisk. Det diskuteres efterfølgende, hvad denne analyse betyder for vores opfattelse af FCI-testen.

Slutteligt gives der perspektiver på, hvordan specialets undersøgelser kan anvendes fremadrettet og der konkluderes på, hvilke resultater vi har set.

## 2 Læringsteori

I det følgende gives et overblik over forskellige aspekter af læringsteori, hvilket er væsentligt i forhold til dette projekts formål af flere forskellige årsager.

Et væsentligt forskningsspørgsmål i denne afhandling handler om at undersøge hvordan en ændring af FCI's evalueringsformat influerer studerendes besvarelser. Først og fremmest må man gøre sig klart, at evaluering er tæt knyttet sammen med læringsteori. Et givet læringsteoretisk synspunkt giver sig udslag i et bestemt blik på evaluering, hvorfor det er nødvendigt at begrunde den teori omkring evaluering, der ligger til grund for ændringen af evalueringsformatet, læringsteoretisk.

Derudover er det nødvendigt med en læringsteoretisk ramme for at analysere og forstå, hvordan de studerendes svar ændres i gentestningen af FCI-spørgsmålene.

### 2.1 Hvad er læringsteori?

Spørgsmålet om hvad læringsteori er, er på en gang let og svært at besvare. Let, fordi læringsteori på den ene side blot er en teoretisk forståelse af hvordan læring finder sted, og svært, fordi denne teoretiske forståelse er en yderst kompleks størrelse. Når jeg i det følgende vil gennemgå forskellige teoretiske standpunkter er der altså ikke tale om en udvikling mod en mere og mere korrekt teori, men snarere om en række af teorier, der kan afdække forskellige aspekter af læringens natur, men som også alle har deres begrænsninger.

At kunne opbygge en teori for, hvordan mennesket lærer og erkender er i høj grad afhængigt af at kunne definere hvad der forstås ved læring og viden. Og her er situationen mere speget end som så fordi der ikke er tale om umiddelbart definerbare begreber. Det læringsteoretiske standpunkt man tager er dermed afhængigt af, hvilket syn på viden og læring man anlægger og de forskellige læringsteorier adskilles af deres bagvedliggende epistemologi. (Dolin, 2006, s. 151)

Man kan have en tendens til at tænke på viden som en absolut størrelse, der lagres i hjernen og kan tages frem efter behov. Problemet med denne tankegang er, at viden tilegnes i en læringsproces og produktet og processen er vanskeligt adskillelige. Selve læringsprocessen er en integreret del af det lærte, og dermed er den viden man har også knyttet sammen med den proces, den er tilegnet igennem. Dette er ikke ensbetydende med, at det lærte ophører med at eksistere når læringsprocessen er tilendebragt, men det peger på en forståelse af viden som noget, der er situeret i den kontekst det er lært i.

### 2.2 Behaviorisme

Som et udgangspunkt for at diskutere konstruktivistisk læringsteori kan det være værd at overveje, hvordan man opfatter læring hvis man konsekvent forstår viden som noget der eksisterer i egen ret og som kan overføres fra lærer til elev, således at eleverne opfattes som tomme kar eller blanke tavler, der skal fyldes op med viden.

I denne måske simpleste forståelse ses læring som elevens respons på de stimuli der præsenteres i undervisningen. Læring bliver dermed en ændring af adfærd, og idealet bliver at arbejde fra et naturvidenskabeligt, behavioristisk udgangspunkt, hvor kun det, der objektivt kan måles, er interessant (Dolin, 2006, s. 155). Den behavioristiske læringsteoretiker beskæftiger sig således ikke med de mentale processer, der finder sted i en læringsituation, men kun med udkommet af det.

## 2.3 Konstruktivisme

Konstruktivismen kan forstås som et opgør med den behavioristiske ide om, at viden er en fast, flytbar størrelse, der eksisterer uafhængigt af individet. I den konstruktivistiske forståelse er viden i stedet noget der konstrueres i læringsprocessen. Viden tilegnes altså gennem en aktiv omformning og det bliver dermed interessant at beskæftige sig med de processer, der forårsager denne bearbejdning.

I det følgende inddeles konstruktivismen lidt forsimplet i en individorienteret mentalkonstruktivisme med udgangspunkt i Piaget og en sociokulturelt orienteret socialkonstruktivisme med udgangspunkt i Vygotskys kulturhistoriske skole.

### 2.3.1 Piaget og mentalkonstruktivisme

Som en helt central figur i den klassiske konstruktivisme står schweizeren Jean Piaget (1896-1980), der betragtede sit virke som genetisk epistemologi, altså som læren om, hvordan viden opstår og udvikles.

I sine studier af dette udviklede Piaget et begrebsapparat, der kan beskrive de kognitivt orienterede konstruktivistiske grundideer. Først og fremmest førte hans undersøgelser om, hvordan viden udvikles hos børn ham til at beskrive den kognitive udvikling ved hjælp af 4 såkaldte stadier, der hver især er kendetegnet ved evnen til anvendelse af bestemte tankestrukturer. Således bliver den viden, et barn har mulighed for at tilegne sig, afhængig af dets mentale udvikling. En væsentlig indvending mod dette er dog ideen om, at når først eleven har nået et givet stadium vil de tankestrukturer der følger med være tilgængelige for eleven i alle tænkelige sammenhænge. Dette strider i høj grad mod bevidstheden om, at viden ofte er situeret og ikke nødvendigvis kan overflyttes fra en kontekst til en anden, hvilket Piaget da også selv senere blev bevidst om.

Et andet centralt punkt i den Piagetske konstruktivisme er forestillingen om, at viden organiseres i mentale strukturer som Piaget kalder skemaer. Et skema består af tre ting, nemlig en situation, en aktivitet knyttet til situationen og et forventet resultat af den aktivitet. Når man præsenteres for en ny situation, må den indpasses i de allerede eksisterende skemaer, og stemmer resultatet af behandlingen af situationen er alt i orden. Piaget taler da om, at situationen er blevet assimileret i de eksisterende skemaer. Stemmer forventning og resultat derimod ikke overens vil der opstå en forstyrrelse, en såkaldt perturbation. Da mennesket ifølge Piaget i høj grad søger at opretholde ligevægt, opleves denne perturbation som generende, og man vil søge at ændre i sine oprindelige skemaer, således at de stemmer bedre overens med den forstyrrende situation. Der vil altså ske en skemaændring gennem det Piaget kalder en akkommodationsproces. Hos Piaget bliver læringsprocesser dermed dybest set konstruktion af viden gennem en ændring af allerede eksisterende skemaer.

Uanset om tilpasningen sker via assimilation eller akkommodation er det en vigtig pointe, at læringsprocessen altid involverer elevernes forforståelse i form af de eksisterende skemaer, og en pædagogisk konsekvens af Piagets skemateori bliver derfor, at undervisningen må tilrettelægges så den forholder sig til elevernes eksisterende viden. Samtidig bliver elevaktivitet<sup>1</sup> et centralt pædagogisk omdrejningspunkt. Læreren kan ikke agere tankpasser, der fylder viden i hovedet på de passive elever, men må derimod sikre, at eleverne arbejder aktivt med stoffet, således at de har mulighed for selv at konstruere deres viden.

---

<sup>1</sup> Ikke nødvendigvis konkret eller fysisk aktivitet. Aktiviteten kan lige så vel være mental.

I forhold til denne forståelse af læring er det væsentligt at være bevidst om, at den tager udgangspunkt i, at mennesket nødvendigvis vil søge ligevægten. Igen tillader denne forståelse af læring ikke så godt en forståelse af, at viden er kontekstuel og situeret, i den forstand at ændringen af et skema i en kontekst ikke nødvendigvis medfører en ændring af det samme skemaer i andre kontekster. Der eksisterer netop eksempler på, at en hverdagsforståelse godt kan eksistere sammen med en anden forståelse der aktiveres i formelle, faglige kontekster. En særlig afart af den Piagetske konstruktivisme handler om at forstå læring som *conceptual change*. *Conceptual change*-modeller for læring kan forstås ud fra begreberne *conceptions* og *misconceptions* som udtryk for studerendes forståelse. *Misconceptions* forstås som et begrebs sæt der på den ene eller den anden måde er i modstrid med det veletablerede begrebs sæt som underviseren ønsker at den studerende skal lære. Da disse *misconceptions* godt kan bygge på hverdagsforståelser og derfor være fornuftige og meningsfulde i egen ret foretrækker nogle teoretikere termer som *alternative conceptions* i stedet. Som i den Piagetske konstruktivisme er rationalet i *conceptual change*-modeller for læringsprocesser at individet konstruerer viden ud fra egne tidligere erfaringer og at læring derfor finder sted ved at *misconceptions* konfronteres og individet præsenteres for en uoverensstemmelse mellem den eksisterende *misconception* og verden: "*conceptual change is the restructuring of an individual's prior conceptions so that they are aligned with widely accepted scientific viewpoints*" (Taasoobshirazi & Sinatra, 2011, s. 902). De finere detaljer omkring, hvordan og hvorfor denne *conceptual change* finder sted (eller ikke finder sted) vil vi ikke komme nærmere ind på her.

### 2.3.2 Vygotsky og socialkonstruktivisme

Hvor Piagets konstruktivisme er meget individorienteret, betoner socialkonstruktivismen betydningen af den sociale kontekst som viden konstrueres i. Viden kan ikke flyttes mellem forskellige kontekster uden at ændres.

Centralt i socialkonstruktivismen står den kulturhistoriske skole med den russiske psykolog Lev Vygotsky (1896-1934), der studerede de mentale processer i et kulturelt og historisk perspektiv. Vygotsky beskæftigede sig også med udviklingen af mentale strukturer, men ifølge ham sker transformationen fra lavere til højere mentale processer ved barnets interaktion med voksne, der påvirker dem både sprogligt og kulturelt. Kultur forstås her som den samling af forestillinger, vurderinger, viden og artefakter som vi tilegner os gennem samspil med omgivelserne. (Säljö, 2003, s. 31)

Udviklingen af de mentale strukturer, og altså dermed læringsprocessen, bliver derfor hos Vygotsky tilegnelsen og evnen til anvendelsen af kulturens sproglige og tekniske hjælpemidler. Kulturen eksisterer som noget interpersonelt som individet gør til sit eget ved at optage dens elementer – mennesket er internaliseret kultur. Denne internaliseringsproces kan forstås som en mediering af individet med omgivelserne hvilket sker ved hjælp af de kulturelle artefakter – vores tænkning er altså farvet af vores kulturs redskaber (Säljö, 2003, s. 87).

Ved artefakter forstås hjælpemidler, der enten kan være konkret materielle, som lommeregneren eller computeren, eller abstrakte og symbolske, som for eksempel sprog og tal, som anvendes i samspil med omverdenen. Disse artefakter er bærere af kulturen og påvirker tankegange og -mønstre hos individet der anvender dem. Artefakterne er således ikke at betragte udelukkende som døde, fysiske genstande. Når vi interagerer med artefaktet interagerer vi også med den menneskelige viden og indsigt som er indbygget i artefakterne, som for eksempel lommeregneren der bærer viden om for eksempel regningsarter og tallet 0,

og som har en betydning for, hvordan man kan løse intellektuelle problemstillinger og dermed ændrer vores mentale aktivitet. Dette betyder også, at artefakterne kun er meningsfulde i en bestemt praksis hvor de bærer en kulturel betydning. Uden for denne praksis mister artefakterne deres mening (Säljö, 2003).

I den socialkonstruktivistiske forståelse af læring spiller sproget som symbolsk artefakt en helt særlig medierende rolle. Således er sprog og tænkning tæt vævet sammen og ikke umiddelbart til at adskille. Man tænker ikke først og formulerer herefter tanken sprogligt. De to processer foregår simultant. Dermed er menneskets forståelse af verden bundet af at skulle gå via det redskab som sproget udgør. Det betyder også, at sproget spiller en central rolle for meningsskabelsen hos eleven i en læringsproces – betydningen skabes gennem sproget. Sproget bruges således til gennem dialog at mediere den personlige tolkning og omverdenens etablerede forståelse af et fænomen (Dolin, 2006).

Dette betyder desuden, at forståelsen i det sociokulturelle perspektiv ses som diskursiv i den forstand at forståelsen er bundet af en måde at anvende sproget på, som er tilpasset traditioner, behov og kultur inden for det fagfelt i relation til hvilket læringen finder sted. Denne forståelse af, at det at lære at anvende sproget er en central del af en læringsproces, kan populært opsummeres som *"learning science is learning to talk science"* (Jay L. Lemke, citeret i (Dolin & Krogh, 2011, s. 65).

Væsentligt i en sociokulturel forståelse af læring står også en forståelse af læringens situerede karakter som det ses hos (Lave & Wenger, 1991). Det lærte er bundet af den sociale og kulturelle kontekst det er tillært i og denne sammenknytning er dobbeltsidig: konteksten influerer ikke kun, hvad der læres, men er også i sig selv bærer af det lærte. Et eksempel på dette ses hos (Säljö, 2003) der refererer til en undersøgelse af brasilianske børn, der sælger kokosnødder på gaden. Deres evne til at udføre multiplikationer er forskellig afhængig af, om den testes relateret til det daglige salg af kokosnødder eller om den testes i mere formaliserede prøvesituationer. Dette understreger den sociokulturelle pointe om, at der ikke eksisterer neutrale kontekster – den viden, individet besidder eksisterer situeret.

Slutteligt er det værd at overveje, hvilke pædagogiske konsekvenser det sociokulturelle syn på læring har. Centralt i disse overvejelser står Vygotskys begreb om zonen for nærmeste udvikling, der betegner forskellen mellem det aktuelle udviklingsniveau og det potentielle udviklingsniveau som det kan opnås gennem problemløsning med hjælp fra lærere eller dygtigere kammerater. Dette begreb får den pædagogiske konsekvens at den gode underviser tilrettelægger læringssituationer der hverken er for trivielle eller for vanskelige for eleven – med andre ord skal underviseren udføre stilladsering i den forstand at han skal understøtte elevens læringsproces netop så meget som det er nødvendigt (Dolin, 2006).

## 2.4 Opsummering og sammenligning af Piaget og Vygotsky

Vi har altså set en skelnen mellem et mentalistisk og et sociokulturelt perspektiv på læring repræsenteret primært, men ikke udelukkende, ved henholdsvis Piagets og Vygotskys arbejde. Noget skarpt kan forskellen på disse to perspektiver opsummeres ved hjælp af modsætningsparrene i nedenstående tabel:

	MENTALISTISK perspektiv	SOCIOKULTURELT perspektiv
Læringens/udviklingens endemål	Evne til formal-logisk tænkning	Beherskelse af kulturelle symboler og strukturer
Pædagogiske grundspørgsmål	Hvordan udvikler vi formal-logiske evner?	Hvordan tilegner vi os vores kultur?
Læringssyn	Individuelt, kognitivt	Socialt, kulturhistorisk
Viden ses som	Kognitive skemaer	Betydningskabelse formet af kulturen
Vidensudvikling	Fra individuel til social	Fra social til individuel
Tilgang til verden	Logisk-deduktiv/ abstrakt-analytisk	Narrativ/sproglig
Pædagogisk praksis	Hensyntagen til individuel videnskonstruktion	Opbygning af sociale processer

Figur 1 Forskellen på et mentalistisk og et sociokulturelt læringsteoretisk perspektiv (Dolin, 2006, s. 168)

Denne helt grundlæggende introduktion til forskellige læringsteoretiske standpunkter giver nu blandt andet mulighed for at diskutere disse teoriers betydning for evalueringsbegrebet.



### 3 Evaluering

I dette afsnit afdækkes forskellige aspekter af evalueringsbegrebet med det formål at blive i stand til at definere to evalueringsparadigmer, et traditionelt og et sociokulturelt, som den udførte empiri tager afsæt i.

Allervigtigst er det at pointere, at man, når man diskuterer evaluering, må starte med at gøre sig klart hvad man vil evaluere. I en didaktisk kontekst vil det som oftest være enten læring eller undervisning. Jeg vil udelukkende beskæftige mig med evaluering af læring. Helt kortfattet kan begrebet evaluering altså i denne opgaves kontekst forstås som et forsøge på at måle og opnå viden omkring hvilke læringsprocesser der har fundet sted hos en given elev på et givet tidspunkt. Dette kan ske gennem eksaminer, tests, feedback på afleveringsopgaver, porteføljeevaluering og så videre. Evalueringsbegrebet er dog mere komplekst end denne simple definition indfanger og at få udforsket disse kompleksiteter er formålet med det følgende.

#### 3.1 Begrundelser for evaluering

Det er ikke naturgivet, at man nødvendigvis skal udføre en eksplicit måling eller bedømmelse af en elevs læring. I det følgende beskrives nogle af de begrundelser der hyppigt anvendes for at foretage evalueringen.

For at kunne gøre dette er det nødvendigt at starte med at gøre sig klart, at evaluering alt efter formål kan klassificeres som intern eller ekstern. Intern evaluering bruges inden for den institution hvor læringsprocessen har fundet sted, hvorimod ekstern evaluering rækker ud over dette – eksempelvis en studentereksamen, der eksternt bruges til for eksempel at søge om optagelse på en videregående uddannelse.

Formålet med formaliseret, ekstern evaluering har ifølge (Gipps, 1999) oprindeligt været dels udvælgelse af de bedst egnede kandidater til offentlige embeder, dels certificering, altså garantien for, at eksaminander, der har bestået eksamen har kvalifikationer til at bestride en given stilling. Ekstern evaluering med disse begrundelser går så langt tilbage som til Handynastiet (206 f.kr. til 220 e.kr.). Noget senere begyndte man også at anvende evalueringen til at differentiere eleverne med henblik på fordeling af studiepladser på videregående uddannelser. Det er stadig et væsentligt argument for at anvende evaluering med karaktergivning på ungdomsuddannelserne, da det regulerer adgangen til de videregående uddannelser.

At bruge denne formaliserede, eksterne evaluering til udvælgelse og certificering til studiepladser og professioner kan ses som et forsøg på at stille kandidaterne lige og på den måde undgå nepotisme. Man må dog ikke være blind for, at de evalueringsformer man bruger ikke er ufejlbarlige i forhold til at stille de forskellige elever lige. (Gipps, 1999) påpeger således, at der er en tendens til, at uddannelsessystemer reproducerer sig selv i den forstand, at elevtyper, der har mange fællestræk med dem, der har udformet evalueringen, har en tendens til at klare evalueringen bedre end elever med færre fællestræk. Denne tendens er i forhold til specialet særlig interessant ud fra et kønsperspektiv – evaluering i et system eller en uddannelse der er kønsskæv, kan have en tendens til at indeholde en kønsbias og dermed stille det underrepræsenterede køn svagere i forhold til at klare evalueringen godt. I forhold til den interne evaluering, som i sagens natur jo ikke bruges til fordeling og certificering til eksterne studiepladser og jobs, er begrundelserne nogle andre.

Den interne evaluering bruges blandt andet til at give læreren et billede af, hvilke elever der ikke får det ønskede læringsudbytte af undervisningen og derfor kan have brug for specialundervisning eller lignende. Ud fra et idealistisk synspunkt kan den interne evaluering dermed bruges til at give eleverne et bedre udbytte af undervisningen end de ville få uden evalueringen. Desuden giver den interne evaluering eleven information om sit eget udbytte af undervisningen, således at hun kan ændre sin indsats i en mere hensigtsmæssig retning. Opsummerende ser vi altså, at evaluering er et udtryk for kontrol i forhold til differentiering af elever og fordeling af jobs og studiepladser. Dermed har man, når man evaluerer elever, et særligt ansvar for, at evalueringen er fair og stiller forskellige elevtyper lige.

### 3.2 Formativ og summativ evaluering

Grundlæggende i forhold til at karakterisere forskellige typer af evaluering står en skelnen mellem formativ og summativ evaluering. Hvad disse begreber dækker over forstås måske bedst ved at opfatte dem som modsætningspar, selvom man nok snarere må betragte de kategorier, der defineres herved, som yderpunkterne af et kontinuum af evalueringstyper. Summativ evaluering bruges til at vurdere og dokumentere et standpunkt, hvor formativ evaluering bruges til fremadrettet at forbedre læringen ved at evaluere en arbejdsproces. Summativ evaluering bliver således produktorienteret, bagudrettet og bruges typisk ved afslutningen på et forløb, hvor formativ evaluering er procesorienteret, peger frem i tiden og foregår undervejs i læringsprocessen (Winsløw, 2006).

Herudover kan det være relevant at skelne mellem kvantitative evalueringsformater, hvor vurderingen er en karakter, en procentsats eller lignende og kvalitativ evaluering, hvor vurderingen for eksempel er en udtalelse (Winsløw, 2006).

### 3.3 Evalueringens validitet og reliabilitet

Inden jeg vil redegøre for, hvordan forskellige evalueringssyn udspringer af forskellige læringssyn, er det nødvendigt at introducere to begreber, der kan bruges til at karakterisere forskellige evalueringstyper, nemlig evalueringens reliabilitet og validitet. Særligt validitetsbegrebet og forståelsen heraf er komplekst, så det følgende skal ses som en temmelig grundlæggende introduktion til begreberne.

Med evalueringens reliabilitet forstås en vurdering af, hvor pålideligt resultatet af evalueringen er. Evalueringsformer der giver samme resultat uafhængigt af ydre omstændigheder siges at have en høj grad af reliabilitet. Således er multiple choice-tests et eksempel på en evalueringsform med høj reliabilitet. Resultatet af testen kan reproducere blot ved at gentage den, resultatet bliver det samme uanset hvem der bedømmer testen og forskellige elevers resultater er lette at sammenligne indbyrdes – så at sige er evalueringens måleusikkerhed minimeret (Winsløw, 2006).

Hvor reliabilitetsbegrebet nogenlunde entydigt lader sig definere er validitetsbegrebet mere sammensat. Helt overordnet er et evalueringsformats validitet et udtryk for, i hvor høj grad formatet faktisk tester det der er intenderet, at det skal teste. Således hedder det med et citat af Wynne Harlen i (Dolin & Krogh, 2011, s. 31): *"Validity means how well what is being assessed corresponds with the behaviour or learning outcome that is intended should be assessed."*

Validitetsbegrebet dækker dog over så forskelligartede spørgsmål som for eksempel hvorvidt evalueringen er udformet, så den indfanger de relevante kompetencer, om evalueringen sker på en måde der tillader tilstrækkelig udfoldelse og udmåling af disse kompetencer og om andre metodevalg ville føre til samme resultat (Dolin & Krogh, 2011, s. 32). En mere nuanceret forståelse af validitetsbegrebets indhold er kompleks og omfattende og ikke nødvendigvis den samme inden for forskellige videnskabelige traditioner.

Tilgangen til validitetsbegrebet i dette speciale bliver derfor overordnet ovenstående meget generelle definition. I kapitel 5.4 diskuteres validiteten af FCI som den er etableret i (Hestenes et al., 1992). Dette validitetsbegreb bliver herefter udgangspunktet for senere at diskutere validiteten af FCI set i lyset af den udførte empiri, således at diskussionen ikke forplumres af eventuelle forskelle i forståelsen af validitetsbegrebet.

En pointe hos (Gipps, 1999) er, at der er forskel på, hvor højt reliabilitet og validitet vægtes afhængig af formålet med evalueringen. I ekstern evaluering er reliabiliteten således væsentlig, fordi resultaterne skal være indbyrdes sammenlignelige, mens validiteten bliver vigtigere i evaluering der har mere diagnostisk eller formativ karakter. I designet af en evaluering bliver det således væsentligt at finde en passende balance mellem reliabilitet og validitet som er konsistent med evalueringens formål og karakter.

### 3.4 Evalueringsparadigmer og læringsteori

Forståelsen af, hvad viden er og hvordan vi tilegner os den hænger naturligvis sammen med forståelsen af, hvordan vi kan måle tilegnelsen af denne viden. Det betyder, at forskellige læringsteorier med tilhørende epistemologier medfører forskellige syn på evaluering – hvad vi kan måle og hvordan vi skal gøre det.

Ud fra et behavioristisk læringsteoretisk synspunkt er viden noget der eksisterer i egen ret og som en fast, uforanderlig størrelse kan overføres fra læreren til eleven. Det interessante for behavioristen er ikke læringsprocesserne hos eleven, men den ændring af adfærd som læringsprocessen resulterer i. Dette hænger tæt sammen med en opfattelse af, at man objektivt kan måle, i hvor høj grad denne overførsel er lykkedes. Idealet er naturvidenskabeligt og man kan tale om, at tilgangen til evaluering er psykometrisk. Inden for denne tilgang til evaluering vil man typisk være interesseret i, at evalueringen har en så høj grad af reliabilitet som muligt, da forbilledet er naturvidenskabernes objektive målinger. Dette kan læringssyn kan give et temmelig curriculumsfokuseret blik på evalueringen. Evalueringen skal tjekke, hvor stor en del af pensum, opfattet som en "*distinct body of information*" der er overført til eleven (Gipps, 1999). I en skoletradition, hvor dannelsesbegrebet er centralt, kan dette blik på evaluering derfor være problematisk.

Ud fra et konstruktivistisk og måske særligt socialkonstruktivistisk læringsteoretisk synspunkt er denne psykometriske tilgang til evaluering ikke meningsfuld. For konstruktivisten eksisterer viden ikke som en selvstændig størrelse, men konstrueres hos den enkelte elev. Processen, hvorved viden tilegnes, kan ikke adskilles fra resultatet, den opnåede viden, og følgelig kan man ikke gøre sig forhåbninger om at kunne måle denne viden adskilt fra de sociale og kontekstuelle forhold som elevens videnskonsstruktion er situeret i. I den socialkonstruktivistiske læringsteori bliver den sociale kontekst særligt vigtig og evaluering må derfor finde sted i samme sociale kontekst som den viden, man ønsker at evaluere, er blevet konstrueret i. En typisk evalueringssituation hvor en elev bliver bedt om for eksempel at regne fysikopgaver alene i en gymnastiksal er derfor uhensigtsmæssig med dette læringsteoretiske udgangspunkt. Da den læring, man ønsker at evaluere, typisk har fundet sted i en kontekst præget af muligheden for social interaktion med både andre elever og undervisere, skal evalueringsformen for at afspejle det læringsteoretiske udgangspunkt situeres i en social kontekst der efterligner den, læringen er sket i, altså med mulighed for samarbejde.

Med til denne ide om, at evalueringen må placeres i en kontekst med mulighed for socialt samspil hører også ideen om, at eleven i evalueringssituationen skal placeres i en stilladseret situation, der giver mulighed for, at eleven kan nå sine nærmeste udviklingszoner.

Evalueringen skal således lede efter elevens best performance snarere end typical performance (Gipps, 1999, s. 375). Således hedder det i (Dolin & Krogh, 2011, s. 29) at i et sociokulturelt evalueringssyn *"er sigtet ikke at tilvejebringe objektiv og reproducerbar viden (hvilket anses for omsonst), men at skaffe valid information, som er meningsfuld og potentielt brugbar for de direkte aktører med kendskab til situationen."*

En central pointe hos Vygotsky er, at vi lærer ved hjælp af kulturens artefakter, både helt konkrete redskaber, men også symbolske og sproglige. For at evalueringen kan være i overensstemmelse med læringsprocessen må man give eleverne mulighed for også i evalueringssituationen at anvende disse artefakter (Gipps, 1999, s. 375).

Man kan altså tale om grundlæggende to forskellige syn på evaluering, nemlig et traditionelt, positivistisk blik med rod i et mentalorienteret læringsteoretisk udgangspunkt og et sociokulturelt blik med rod i den socialkonstruktivistiske læringsteori. Således skelnes der i (Dolin & Krogh, 2011) mellem et postpositivistisk og et sociokulturelt evalueringssyn. Skematisk kan dette illustreres ved hjælp af nogle modsætningspar, der giver et billede af forskellen på de to blik på evalueringen, selvom disse modsætningspar naturligvis ikke udgør en udtømmende definition. Et sådant skema ses herunder:

### Traditionelt paradigme

- Ikke-interaktiv
- Ikke-kollaborativ
- Statisk
- Produktorienteret
- Begrænset brug af artefakter
- Løsrevet fra situeret og autentisk praksis

### Sociokulturelt paradigme

- Interaktiv
- Kollaborativ
- Dynamisk
- Procesorienteret
- Udstrakt brug af artefakter
- Indlejret i situeret og autentisk praksis

Figur 2 Forskellen på evalueringssyn (Dolin & Krogh, 2011, s. 28)

I forhold til at karakterisere forskellige evalueringssyn efter denne inddeling er det klart, at der ikke er tale om en binær inddeling, hvor evalueringen nødvendigvis falder i en af de to kategorier, men at der derimod snarere er tale om yderpunkterne i et kontinuum og at forskellige evalueringssyn kan være mere eller mindre sociokulturelt eller traditionelt orienterede.

Vurderingen af de forskellige evalueringssyns gyldighed afhænger derfor i høj grad af det epistemologiske og læringsteoretiske udgangspunkt man tager. Opstillingen af specialets andet forskningsspørgsmål forudsætter derfor et blik på læring og erkendelse der er socialkonstruktivistisk orienteret og som altså har som udgangspunkt, at viden er situeret i sociale kontekster og at dialog og anvendelse af sproget som artefakt til at konstruere forståelse med er centralt. I (Gipps, 1999) pointeres det desuden, at sociokulturelt orienterede evalueringssyn ikke altid lever op til traditionelle kriterier om reliabilitet. Dette medfører nogle metodiske vanskeligheder i relationen til sammenligningen af forskellige evalueringssyn, som vi skal vende tilbage til i kapitel 8.9.

## 4 Køn

Med køn som et underlæggende tema i dette speciale vil jeg forsøge at placere specialet i relation til andre studier, der omhandler fysik og køn.

(Danielsson, 2009) gennemgår i et literature review den eksisterende litteratur der omhandler krydsfeltet mellem physics education research og kønsstudier. Denne litteratur inddeles i forskellige kategorier, blandt andet ved at overveje, om begrebet køn diskuteres og problematiseres eller blot opfattes som et synonym for kvinder og mænd. En væsentlig kategori er Comparisons of Man and Woman Students. Ifølge Danielsson har studierne i denne kategori en tendens til at " ...treat gender as a stable category and focus on the differences between the genders rather than on the variations within the genders." Her opfattes kønnene i højere grad som fast definerede, uforanderlige størrelser end som sociale konstruktioner, og man opfatter køn om et primært karakteristikum for forsøgspersonerne. Man er altså primært sit køn, sekundært fysiker med de identiteter det indebærer. Når dette ydermere ofte kombineres med en tilgang til køn og naturvidenskab der handler om "*the women question in science*" snarere end "*the science question in feminism*", hvorved kvinderne snarere end videnskaben opfattes som problemet (Danielsson, 2009, s. 31), kan man have en tendens til at konstruere to typer af fysikstuderende: mænd med en naturlig interesse og flair for faget og kvinder med en lavere faglig selvtillid og større faglige problemer.

Dette giver selvsagt et på mange måder unuanceret billede, og det kan derfor være hensigtsmæssigt at overveje tilgange der i højere grad er følsomt overfor forskelle internt i kønnene, således at der åbnes op for en bredere vifte af identiteter som henholdsvis mandlig og kvindelig fysiker. Dette kan for eksempel ske ved at betragte kønsidentiteter som sociale konstruktioner der skabes i det sociale praksisfællesskab, der udgøres af fysikfaget.

Den del af specialet, der relaterer til det første forskningsspørgsmål, anlægger i høj grad det blik på køn, der handler om at sammenligne mandlige og kvindelige studerende.

Udgangspunktet bliver her at undersøge, hvordan ens FCI-score er relateret til ens køn som det primært definerende karakteristikum. Dette er, som anført ovenfor, et noget unuanceret blik på køn. Når det alligevel gøres skyldes det primært, at det er denne tilgang til køn der ses i eksisterende litteratur (for eksempel (Birch & Walet, 2012) og (Dockett & Heller, 2008)) om FCI-testen i et kønsperspektiv. For at kunne lade undersøgelserne tage udgangspunkt i denne litteratur er det hensigtsmæssigt at tage det samme kønsteoretiske udgangspunkt. Tilgangen til køn nuanceres til gengæld en smule i specialets anden del, hvor den kvalitative analyse tillader, at vi beskæftiger os med et bredere spektrum af både mandlige og kvindelige studerende, der ikke nødvendigvis er udvalgt som typiske repræsentanter for deres køn. Opsummerende kan det altså siges, at kønsbegrebet i dette speciale grundlæggende er forholdsvis traditionelt, hvilket skyldes et ønske om at kunne arbejde ud fra eksisterende litteratur om køn og FCI, men at det samtidig er værd at være bevidst om, at denne tilgang kan problematiseres ved at betragte køn som noget, der også er socialt konstrueret.

## 5 Force Concept Inventory

I det følgende afsnit gives først en introduktion til The Force Concept Inventory (FCI). Herefter beskrives tidligere fremsatte kritikpunkter angående FCI's anvendelighed og diskussion af, hvordan FCI-resultaterne skal fortolkes og slutteligt gives en analyse af, hvilken lærings- og evalueringsteoretisk tradition FCI er udviklet i.

### 5.1 Hvad er FCI?

The Force Concept Inventory, forkortet FCI, er en test af studerendes konceptuelle forståelse af klassisk Newtonsk mekanik. FCI er udformet som en multiple choice-test og består af 30 spørgsmål, hver med 5 forskellige svarmuligheder. Testen er udelukkende konceptuel og der optræder altså hverken talværdier eller udregninger i den. Testen blev første gang publiceret i en artikel i *The Physics Teacher* i 1992 (Hestenes, Wells & Swackhamer, 1992), og er en videreudvikling af den såkaldte Mechanics Diagnostics-test (MD), som blev publiceret i *American Journal of Physics* i 1985 (Halloun & Hestenes, 1985a). Såvel FCI som MD baserer sig på forfatterens arbejde med at undersøge og klassificere typiske misconceptions omkring klassisk mekanik hos amerikanske college-studerende. I 1995 publiceredes en revideret udgave af FCI<sup>2</sup> og siden er FCI blevet et helt centralt instrument til evaluering af fysikstuderendes konceptuelle forståelse, hvilket både afspejles i dens store udbredelse på diverse uddannelsesinstitutioner og i de anseelige mængder litteratur der er publiceret om FCI (Rebsdorf, 2006).

FCI's fysikfaglige udgangspunkt er, at den centrale udfordring man skal overkomme i forhold til at opnå en korrekt Newtonsk tankegang er tilegnelsen af kraftbegrebet. *"Without this concept the rest of mechanics is useless, if not meaningless,"* som det hedder i (Hestenes et al., 1992). Jeg vil ikke problematisere denne opfattelse af den klassiske mekanik yderligere, men blot pointere, at det er denne tankegang, der ligger til grund for FCI – at klassisk mekanik forstås via en forståelse af kraftbegrebet.

I (Hestenes et al., 1992) inddeles den klassiske mekanik ud fra dette præmis i 6 såkaldte konceptuelle dimensioner, nemlig kinematik, Newtons 1., 2. og 3. lov, superposition af kræfter og forskellige typer af kræfter. Det er forståelsen af disse 6 dimensioner der ønskes undersøgt via FCI.

I (Halloun & Hestenes, 1985b) er det undersøgt hvilke misconceptions fysikstuderende typisk har omkring hver af disse konceptuelle dimensioner. I FCI er disse misconceptions anvendt som svarmuligheder i spørgsmål vedrørende den tilsvarende konceptuelle dimension som det der i (Hestenes & Halloun, 1995) kaldes "powerful distracters". Rationalet er altså, at alle forkerte svarmuligheder i et givet spørgsmål i FCI er faktiske misconceptions som der er empirisk belæg for at tro eksisterer hos studerende. Ud fra et givet forkert svar i FCI kan man altså ved hjælp af en tabel i (Hestenes et al., 1992) direkte aflæse, hvilken misconception dette forkerte svar må formodes at dække over.

---

<sup>2</sup> Det er i litteraturen udbredt at anvende litteratur der tager udgangspunkt i 1992-udgaven af FCI-testen til at beskrive 1995-udgaven med da forskellene på de to udgaver er beskedne. Dette er således også gjort her.

## 5.2 FCI-gain

En væsentlig pointe omkring anvendelsen af FCI er, at den gerne gives både i starten og slutningen af et undervisningsforløb, altså som såkaldte præ- og posttests. Dette giver mulighed for at udregne hver enkelt studerendes gain<sup>3</sup> som et mål for, hvor meget den konceptuelle forståelse ændres i løbet af undervisningsforløbet. FCI kan læses som en kritik af det der hos (Hestenes et al., 1992) betegnes som konventionel mekanikundervisning, der ikke er en effektiv metode til at sikre en konceptuel mekanisk forståelse hos de studerende i form af et tilfredsstillende FCI-gain. Som det hedder i (Hestenes et al., 1992) kan "*pedagogy (...)* *make a difference*" og en helt grundlæggende pointe i forhold til brugen af FCI bliver, at gainet kan bruges som et mål for, hvilke undervisningsmetoder, man skal vælge for at sikre en mekanikundervisning, der øger de studerendes konceptuelle forståelse. Formålet med dette speciale er dog ikke at beskæftige sig med mekanikundervisning, men derimod med selve FCI-testen ud fra dels et kønsperspektiv og dels et evalueringsteoretisk perspektiv. Derfor beskæftiger specialet sig ikke med FCI-gain, men derimod udelukkende med præscorer.

## 5.3 FCI's anvendelsesmuligheder

I (Hestenes et al., 1992) gives der ikke noget entydigt svar på, hvordan det er hensigten, at FCI skal anvendes – tværtimod åbnes der op for et relativt bredt spektrum af anvendelsesmuligheder. Således hedder det, at FCI-testen er inkluderet i artiklen "*for teachers to use any way they see fit*" og de fleste døre holdes dermed åbne. Implementeringen af testen på forskellige uddannelsesinstitutioner er ikke en umiddelbar del af artiklens interessefelt. Dog angives der alligevel tre forskelligartede anvendelsesmuligheder, nemlig at testen kan bruges som et diagnostisk værktøj, til at evaluere undervisning med og som såkaldt placement exam. der bruges til at afgøre, hvilke studerende, der kunne tænkes at have problemer med givne kurser der forudsætter et kendskab til klassisk mekanik. Det udtrykkes, at testen muligvis ikke bør stå alene hvis den anvendes til dette sidste formål. Det ses altså, at testen både tænkes brugt som et værktøj inden for forskning i fysikdidaktik og med undervisningsformål for øje, idet undervisere kan bruge testen til at opnå indsigt i misconceptions hos studerende.

## 5.4 FCI's validitet og reliabilitet

I det følgende gennemgås, hvordan FCI-testens validitet og reliabilitet er etableret i (Hestenes et al., 1992). Som nævnt i kapitel 3.3 gøres dette uden at validitets- og reliabilitetsbegreberne, der anvendes her, problematiseres yderligere.

Reliabiliteten (som dette begreb er defineret i kapitel 3.3) af FCI må forventes at være høj. Multiple choice-formatet gør, at resultaterne er reproducerbare og let sammenlignelige forudsat at ydre rammer som for eksempel den tid, der er til rådighed til at tage testen, holdes konstante.

Diskussionen omkring validitet og reliabilitet udfoldes kun i begrænset grad i (Hestenes et al., 1992). Det anføres dog, at der er gennemført to interviewsessions med henholdsvis 16 og 20 studerende omkring deres besvarelser af FCI-spørgsmålene. Herudover refereres i stedet til etableringen af validiteten og reliabiliteten af MD-testen.

I (Halloun & Hestenes, 1985a) er validiteten af MD-testen etableret på flere forskellige måder: fysikprofessorer og ældre studerende er kommet med forslag til forbedringer, en test af ældre

---

<sup>3</sup> Det normaliserede gain defineres som  $\frac{\text{postscore} - \text{præscore}}{100\% - \text{præscore}}$  og fortæller altså, hvor stor en andel af den maksimalt mulige forbedring den faktisk opnåede forbedring udgør.

studerende viste, at de var enige om det korrekte svar på de forskellige spørgsmål og studerende med høje fysikkarakterer blev testet og deres besvarelser blev gennemgået uden at der blev fundet tegn på misforståelser der bundede i formuleringen af spørgsmålene. Slutteligt blev 22 nystartede studerende interviewet for at tjekke, om de forstod spørgsmålene og de forskellige svar. Validiteten er altså veletableret i den forstand at der er bred enighed om, at spørgsmålene er forståelige, utvetydige og relevante i forhold til det faglige område Newtonsk mekanik, og at FCI-testen i den forstand faktisk er en test af konceptuel forståelse af kraftbegrebet i den klassiske mekanik.

Det hedder desuden i (Halloun & Hestenes, 1985a), at reliabiliteten af MD-testen er kontrolleret ved opfølgende interviews med studerende der er blevet testet ved hjælp af MD-testen. Det viste sig, at de studerende i stort set alle tilfælde gav de samme svar i interviewene som de havde givet i MD-testen og at deres overbevisning om korrektheden af disse svar var ganske fast. I den forståelse af reliabilitet og validitet der er introduceret i kapitel 3.3 er dette dog snarere en kontrol af testens validitet. Denne forståelse genfindes da også i (Hestenes et al., 1992), hvor validiteten af FCI jo netop kontrolleres via interviews. Vi vil altså betragte dette som et andet aspekt af FCI-testens validitet: ifølge (Halloun & Hestenes, 1985a) er FCI-testen valid i den forstand, at de studerendes forståelse ikke er anderledes her end hvis man søger at skaffe sig adgang til den via andre metoder.

I interviewene gennemført i (Hestenes et al., 1992) ses ligeledes et billede af, at de studerende svarer på samme måde her som de gør i FCI. Det viser sig, at der kun sjældent er tilfælde af, at et forkert svar vælges af en studerende som besidder de relevante Newtonske koncepter, mens det sommetider ses at et korrekt svar vælges med ikke-Newtonske begrundelser.

Dette indikerer altså, at FCI giver et retvisende billede af den øvre grænse for de studerendes Newtonske forståelse i den forstand, at der ses en konsistens mellem besvarelserne i FCI og interviewene. Metoden der er anvendt til disse interviews er dog ikke beskrevet nærmere i hverken (Hestenes et al., 1992) eller (Halloun & Hestenes, 1985a), hvilket gør, at det er svært at diskutere nærmere forhold om rækkevidden af disse konklusioner omkring validiteten.

## 5.5 Tidligere kritik af FCI

Et så udbredt værktøj som FCI har naturligt nok været diskuteret i fysikdidaktiklitteraturen. Et væsentligt kontrovers omkring anvendelsen og fortolkningen af FCI er centreret omkring (Huffman & Heller, 1995a) og den efterfølgende besvarelse af kritikken i (Hestenes & Halloun, 1995).

I (Huffman og Heller, 1995a) er der foretaget en såkaldt factor analysis af FCI-data for både high school- og universitetsstuderende. Disse analyser viser, at inddelingen af spørgsmålene i de 6 konceptuelle dimensioner ikke genfindes i den måde, de studerendes korrekte besvarelser grupperes i sammenhængende faktorer på. Grundlæggende ville man forvente, at for eksempel alle spørgsmål omkring gravitation forstås som et udtryk for det samme koncept, og at en studerende der besvarer et af disse spørgsmål korrekt også vil besvare de andre spørgsmål i samme konceptuelle dimension korrekt. At dette ikke er tilfældet indikerer, mener (Huffman og Heller, 1995), at FCI-testen måske ikke så meget måler et samlet kraftkoncept hos de studerende som "bits and pieces" af et sådant kraftkoncept. Dette forsøges forstået læringsteoretisk på flere forskellige måder. Således forstås disse forståelsesbidder som såkaldte phenomenological primitives eller p-prims. Dette er simple mentale strukturer, der aktiveres hos den studerende når en situation skal forklares. Disse p-prims er hverken korrekte eller forkerte i sig selv – den samlede forståelse kan derimod være



korrekt eller forkert afhængig hvilken sammensætning af p-prims der aktiveres (Hammer, 1996, s. 1319).

Desuden påpeges det, at studerendes viden ofte er kontekstuel situeret, således at det er svært at adskille i hvor høj grad FCI-testen måler studerendes forståelse af et koncept som hvorvidt den kontekst konceptet optræder i er genkendelig.

I (Hestenes & Halloun, 1995) besvares kritikken ved at sige, at inddelingen af kraftbegrebet i 6 konceptuelle dimensioner ikke skal ses som en beskrivelse af de studerendes verdensbillede, men snarere som et i fagfællesskabet veletableret spejl at holde denne forståelse op imod for at teste, i hvor høj grad de studerende har tilegnet sig dette Newtonske verdensbillede. Dette er ganske givet korrekt, men rykker ikke nødvendigvis ved konklusionen i (Huffman & Heller, 1995a), nemlig at der er behov for at forstå de studerendes besvarelser af FCI-testen læringsteoretisk fordi analysen peger på, at de studerende ikke udvikler en samlet forståelse af hver enkelt konceptuelle dimension i den Newtonske mekanik på en gang. Der er derimod tale om en meget mere kompleks konstruktion af forståelse hos den studerende – noget (Hestenes & Halloun, 1995) da også åbner op for ved at medgive, at de er enige i, at den ikke-Newtonske tænkens verdensbillede ofte er fragmenteret og usammenhængende. Således hedder det i (Hestenes & Halloun, 1995) at *"Student concepts are often "vague and undifferentiated" and they are "incompatible with Newtonian concepts in most respects"*". Den ikke-Newtonske tænkens verdensbillede er altså usammenhængende og ikke nødvendigvis internt konsistent – men det indeholder sjældent Newtonske elementer.

## 5.6 Lærings- og evalueringsteoretisk analyse af FCI

I (Hestenes et al., 1992) gøres hverken det lærings- eller det evalueringsteoretiske udgangspunkt for FCI-testen klart. Det er derfor temmelig implicit, hvilket syn på læring og viden, der udtrykkes via FCI og formålet med følgende afsnit er at analysere dette.

### 5.6.1 Læringsteoretisk analyse

Den grundlæggende præmis for FCI-testen er, at alle studerende møder klassisk mekanik som institutionaliseret fag i uddannelsesverdenen med en hverdagsforståelse der baserer sig på tidligere dagligdags erfaringer. Med (Hestenes et al., 1992)'s egne ord hedder det således: *"Every student begins physics with a well-established system of commonsense beliefs about how the physical world works derived from years of personal experience. Over the last decade, physics education research has established that these beliefs play a dominant role in introductory physics."* Begrebet misconceptions bruges til at karakterisere disse commonsense beliefs, men det pointeres dog, at begrebet misconceptions grundlæggende er uheldigt, da det ikke nødvendigvis er let at indse, at disse misconceptions er forkerte. De kan sagtens være både plausible og fornuftige for den studerende.

I forlængelse heraf bliver en væsentlig pointe i FCI-regi også, at der hos mange studerende eksisterer et misforhold mellem den intuitive forståelse af, hvordan ting opfører sig, oparbejdet gennem mange års erfaringer, og den Newtonske forståelse. Eksempelvis kan Newtons 1. lov forekomme kontraintuitiv, da alle hverdagserfaringer jo netop fortæller, at objekter, der ikke påvirkes af kræfter før eller siden vil finde hvile, i det man overser den friktion der påvirker objektet.

Det læringsteoretiske standpunkt som det ses hos (Hestenes et al., 1992) kan derfor kortes ned til, at de studerende potentielt har misconceptions, typisk konstrueret ud fra hverdagserfaringer, der skal konfronteres i undervisningen for at den studerende kan lære nyt. Jeg vil derfor mene, at man kan karakterisere læringssynet bag FCI som konstruktivistisk, da videnstilegnelse hos de studerende tager udgangspunkt i allerede eksisterende viden og

conceptual change-inspireret i den forstand, at de studerende er i besiddelse af misconceptions baseret på hverdagserfaringer som den vellykkede undervisning konfronterer og erstatter med videnskabeligt korrekte conceptions. I (Hammer, 1996, s. 1318) argumenteres der således også for, at et tilsvarende lærings syn er udbredt i naturfagsdidaktiske sammenhænge.

Hestenes et al. anvender hyppigt begreberne Newtonsk tænkning og Newtonske tænkere. Således introduceres der i (Hestenes & Halloun, 1995) to særligt bemærkelsesværdige FCI-scoringer: 60% korrekte svar i FCI udgør et såkaldt entry threshold til Newtonsk tænkning og 85% et mastery threshold. Studerende med scoringer under 60% har en fragmenteret og usammenhængende forståelse af mekanik. Studerende med scoringer mellem 60% og 85% har derimod en forståelse af de mest basale aspekter af den Newtonske mekanik. En komplet forståelse indikeres af en score over 85%. Disse tærskler indikerer en læringsteoretisk forståelse af, at det Newtonske verdensbillede hos den studerende udvikler sig akkumulativt, således at det ikke-Newtonske verdensbillede gradvis fortrænges af et Newtonsk verdensbillede, snarere end at et Newtonsk verdensbillede anvendes i nogle kontekster, mens de ikke-Newtonske misforståelser kommer til udtryk i andre sammenhænge.

Som beskrevet i kapitlet om læringsteori er en væsentlig pointe inden for conceptual change-modellen at forskellige conceptions godt kan sameksistere hos den studerende. Uden at det gøres eksplicit i (Hestenes et al., 1992) vil jeg dog mene, at denne forståelse kun i begrænset grad er til stede her. FCI opererer med en misconceptions-taksonomi, om hvilken det hedder, at *"Presence of the misconceptions is suggested by selection of the corresponding Inventory Item"*. Rationalet er altså, at hvis et forkert svar vælges i FCI er den tilhørende misconception til stede – ikke kun i relation til denne specifikke kontekst, men også mere generelt. Således er der hos FCI grundlæggende ikke en forståelse for, at forskellige conceptions kan eksistere samtidig og at den studerendes viden dermed er kontekstafhængig, for eksempel ved, at den studerende tænker ikke-Newtonsk i hverdagssituationer, men Newtonsk i formelle læringssituationer. Dette ses blandt andet i følgende citat fra (Hestenes & Halloun, 1995): *"The FCI probes each conceptual dimension with several questions involving different contexts and viewpoints. A false positive on one of the questions can then be partially compensated by a non-Newtonian choice on another"* Besvares der Newtonsk i en kontekst og ikke-Newtonsk i en anden er det altså ikke så meget et spørgsmål om, at den studerende tænker Newtonsk i nogle kontekster og ikke i andre, som det er en indikation af, at den korrekte besvarelse er valgt ud fra ikke-Newtonske overvejelser. Det påpeges derfor, at man i udgangspunktet ikke må lægge for stor vægt på spørgsmålene enkeltvis, da der kan forekomme falsk positive svar hvor det korrekte svar vælges men ud fra forkerte bevæggrunde eller ved simpelt gætteri. Derimod understreges det, at *"only a true Newtonian generates a consistent pattern of Newtonian choices with an occasional lapse at most. Thus, the Inventory as a whole is a very good detector of Newtonian thinking."* Hestenes et al.'s forståelse, baseret på interviews med studerende, er altså, at der derimod kun i beskeden grad forekommer falsk negative svar, hvor en studerende, der egentlig burde kunne besvare et spørgsmål korrekt alligevel svarer forkert. Dermed bliver FCI-scoringen en øvre grænse for i hvor høj grad testpersonen har tilegnet sig et Newtonsk verdensbillede.

I forhold til at diskutere FCI-testens læringsteoretiske udgangspunkt er det også værd at bide mærke i, hvilket syn på fysikcurriculummet inden for den Newtonske mekanik testen antager. Dette opsummeres i Table 1 i (Hestenes et al., 1992), hvis indhold ikke skal anfægtes. Det er dog interessant, at dette curriculum betragtes som noget, der er konstant på tværs af

kontekst, land og kultur, snarere end som noget, som også er afhængigt af for eksempel læreplaner og undervisningsmål på forskellige uddannelsesinstitutioner i forskellige lande.

### 5.6.2 Evalueringsteoretisk analyse

Jeg vil mene, at det er forholdsvis ukontroversielt at karakterise evalueringsformatet der anvendes i FCI som tilhørende det i kapitel 3.4 definerede traditionelle evalueringsparadigme. Evalueringsformatet har en høj grad af reliabilitet (det er indiskutabelt hvilken score en given studerende har opnået i FCI) og evalueringen skal objektivt måle, i hvor høj grad de studerende har tilegnet sig den etablerede Newtonske tankegang. Evalueringen er altså produktorienteret snarere end procesorienteret. Formatet tillader ingen brug af artefakter og evalueringen foretages løsrevet fra den situerede praksis læringsprocessen har fundet sted i. Der er ingen dialog eller interaktion med den studerende der skal evalueres, ligesom der heller ikke er mulighed for samarbejde elever imellem.

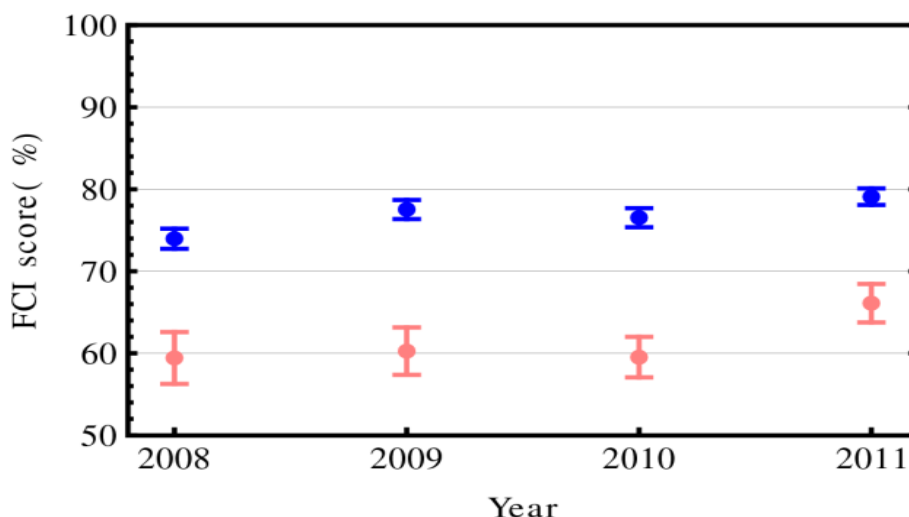
I forhold til alle de parametre, der kan bruges til at karakterisere evalueringsparadigmer med befinder FCI sig altså i udpræget grad i det traditionelle paradigme.

## 6 Kønsforskellen i FCI

### 6.1 Kønsforskellen i udlandet

Det første forskningsspørgsmål i dette speciale har blandt andet til formål at afgøre, om der eksisterer en kønsforskel i scoren på FCI-testen blandt førsteårsstuderende på NBI. Det følgende afsnit fungerer derfor som et overblik over den eksisterende litteratur der beskæftiger sig med kønsforskellen i FCI-resultater i andre lande med det formål senere at kunne vurdere, hvordan data fra NBI passer ind i dette billede.

I (Birch & Walet, 2012) gennemgås 4 års data fra FCI-testen foretaget på University of Manchester i Storbritannien. Testen foretages første gang i den såkaldte Welcome Week inden universitetsundervisningen begynder. Et plot over denne kønsforskel ses nedenfor.



Figur 3 Gennemsnitlig FCI-score for mænd (blå) og kvinder (rød) med error bars der repræsenterer 1 standardfejl. Data er fra prætesten i årene 2008-2011 på University of Manchester (Birch & Walet, 2012)

Der ses altså en kønsforskel, som er statistisk signifikant og som går igen fra år til år. Denne kønsforskel angives til at lægge på mellem 7-17%. Yderligere er det hos (Birch og Walet, 2012) undersøgt, hvordan denne kønsforskel fordeler sig på de 30 spørgsmål. Her er tendensen, at andelen af mandlige studerende der svarer korrekt er højere end andelen af kvindelige studerende på stort set alle 30 spørgsmål. Der er dog forskel på størrelsen af denne kønsforskel fra spørgsmål til spørgsmål. Den største forskel ses for spørgsmål 14 og 23.

I (Docktor & Heller, 2008) analyseres 10 års FCI-data fra University of Minnesota i USA med henblik på at undersøge, om der eksisterer en kønsforskel her. Data er indsamlet på et kursus i "calculus-based physics for scientists & engineers" på første semester, og prætesten er gennemført i første kursusuge. Også her ses samme tendens. Den gennemsnitlige mandlige

FCI-score i prætesten er signifikant højere end den tilsvarende kvindelige score, nemlig  $15,3 \pm 0,5\%$ . Det er interessant, at studerende med mere end tre blanke svar er udeladt af analysen. Dette betyder, at flere blanke svar hos kvinderne ikke kan forklare kønsforskellen her. Omvendt er det karakteristisk, at det kun er 60 studerende, der er udeladt på denne baggrund, hvilket er forsvindende få i forhold til de 5636 studerende, hvis resultater indgår i den endelige analyse. I (Docktor & Heller, 2008) genfindes kønsforskellen i mændenes favør desuden på samtlige 30 FCI-spørgsmål, hvilket stemmer overens med observationerne fra Manchester. Særligt bemærkelsesværdigt er det, at det er netop de samme to spørgsmål (14 og 23) som i Manchester, for hvilke der ses den største kønsforskel. Slutteligt er det værd at bemærke, at (Docktor & Heller, 2008) også har sammenlignet gennemsnitlige eksamenskarakterer i kurset for de to køn. Her ligger mændenes karakterer  $1,5 \pm 0,2\%$  højere end kvindernes. Kønsforskellen fra FCI-resultaterne genfindes altså ikke i eksamenskarakteren.

Opsummerende ses der altså et billede af en kønsforskel i FCI-testen som er bemærkelsesværdigt ens på to forskellige uddannelsesinstitutioner. Dette understreger behovet for dels at undersøge, om den samme kønsforskel ses på Københavns Universitet og dels at søge at forstå, hvad denne kønsforskel skyldes.

## 6.2 Kønsforskellen i FCI – tidligere forklaringsmodeller

Grundlæggende kan man forsøge at forklare den velunderbyggede kønsskævhed i FCI-scorerne på to måder:

Enten er kønsforskellen udtryk for eksisterende forskelle i de to køns konceptuelle mekanikforståelse eller også er der indbygget en kønsbias i testen der gør, at den observerede forskel afspejler en svaghed i testen og ikke en eksisterende kompetenceforskel. Det er dog vigtigt at være bevidst om, at de forskellige forklaringsmodeller ikke nødvendigvis gensidigt udelukker hinanden og at der således kan være tale om flere forskellige aspekter der tilsammen giver et billede af, hvad kønsforskellen skyldes.

Jeg har i litteraturen fundet to studier, der beskæftiger sig med mulige forklaringer på kønsforskellen i FCI som den kommer til udtryk i den enkelte FCI-test og altså ikke som en kønsforskel i FCI-gain, der dermed fortæller om, hvordan kønsforskellen i testresultaterne muligvis kan ændres via undervisning, som det er tilfældet i (Kost, Pollock & Finkelstein, 2009) og (Lorenzo, Crouch, & Mazur, 2006).

I (McCullough, 2004) søges kønsforskellen forklaret som en indbygget bias i testen. Rationalet er her, at der i mange sammenhænge konstrueres et billede af fysik som et maskulint felt, hvilket skaber en kønsulighed i forhold til deltagelse og præstation inden for feltet. Således peges der på, at den kontekst, der præsenteres i spørgsmålene i FCI kan have indflydelse på de studerendes besvarelse. Det er således lettere at besvare spørgsmål hvis kontekst er tæt på ens egen dagligdag og erfaringsområde end spørgsmål, hvis kontekst ikke er velkendt. Disse overvejelser ligger i forlængelse af overvejelserne omkring FCI i kapitel 5.5. *"The test [FCI] includes many stereotypically male contexts such as hockey, rockets and cannonballs"*, skriver McCullough. Der er derfor lavet en revideret udgave af FCI, hvor spørgsmålene har samme fysiske indhold, men hvor spørgsmålene ændres til at omhandle *"stereotypically female contexts such as shopping, cooking, jewelry and stuffed animals"*. Denne reviderede FCI-test blev givet til i alt 150 studerende, hvis besvarelser blev sammenlignet med 162 studerendes besvarelser af den originale udgave af FCI. Overordnet set ses det, at kontekstændringen ikke

betyder noget for de kvindelige studerendes gennemsnitlige score, mens mændenes gennemsnitlige score bliver lavere (fra 33,7% korrekte til 28,5% korrekte). Til gengæld er der ikke noget entydigt mønster i, hvilke spørgsmål der er skyld i denne ændring – *“every combination of relationships imaginable showed up in the test”*. Det konkluderes derfor, at en kontekstændring kan ændre studerendes præstation helt overordnet, men at det ikke er muligt afgøre, præcis hvilke kontekster der betyder hvad for de forskellige grupper af studerende.

I (McCullough, 2004) foregår kontekstændringen på to niveauer: både som en ændring af kontekst fra det stereotyp maskuline til det stereotyp feminine, men også som en ændring til en mere autentisk, hverdagsnær og konkret kontekst, da undersøgelser viser, at kvindelige studerende foretrækker sådanne opgaver. Jeg vil dog mene, at man kan stille spørgsmålstejn ved, om de nye spørgsmål nødvendigvis er mere autentiske og hverdagsnære for de kvindelige studerende. Således ændres for eksempel spørgsmål 12, der i FCI-testen omhandler affyringen af en kanonkugle, til at omhandle en baby, der kaster med en skål med mos. Dette er næppe heller en autentisk og genkendelig kontekst for gruppen af relativt unge collegestuderende kvinder i undersøgelsen.

De medvirkende i undersøgelsen er *“a group of non-physics students in general education classes. [...] Five were English or Literature classes, two were Sociology classes, and one was a math class.”* Disse forsøgspersoner blev valgt, da mænd typisk er overrepræsenterede på fysikholdene og selve det at udgøre en minoritet kan få de kvindelige studerende til at underpræstere. Ved at vælge disse andre hold burde man altså kunne undersøge den eventuelle kontekstinducerede kønsbias i FCI-testen uden denne usikkerhed. Jeg vil dog mene, at man hermed åbner op for nogle andre usikkerheder i forhold til at kunne anvende resultaterne til at forklare den observerede kønsforskel – man måler på forsøgspersoner fra en anden gruppe af studerende end den man faktisk er interesseret i at forklare kønsforskellen i. De gennemsnitlige scorer er specielt for kvindernes vedkommende meget lave – kun godt 20% korrekte i gennemsnit. Da hvert spørgsmål har 5 svarmuligheder vil man kunne opnå sammenlignelige scorer ved ganske enkelt at gætte sig frem. Man kunne forestille sig, at konteksten får en noget anden betydning for studerende, hvis fysikfaglige grundlag er så svagt er de må gætte sig frem, muligvis understøttet af dagligdags erfaringer, end for studerende der i højere grad har mulighed for at besvare spørgsmålene ud fra deres institutionaliserede fysikfaglige viden.

Opsummerende kan vi altså sige, at der ikke er belæg for ud fra (McCullough, 2004) at forklare kønsforskellen ud fra FCI-spørgsmålene som repræsenterende stereotyp maskuline kontekster. Til gengæld ses det, at den kontekst en opgave præsenterer tilsyneladende godt kan have indflydelse på, hvordan studerende besvarer den.

I (Popp, Meltzer & Megowan-Romanowicz, 2011) undersøges FCI ved hjælp af en såkaldt differential item functioning (DIF)-analyse med det formål at afgøre, om der eksisterer en systematisk kønsbias i testen. Formålet med DIF-analysen er at afgøre, om studerende med samme forudsætninger svarer forskelligt på de forskellige dele af FCI afhængigt af, om de er mænd eller kvinder. Dette er gjort ved hjælp af posttestdata fra 4775 high school-studerende fra USA. Analysen er for teknisk til at gengive i rammerne af dette speciale, men formålet er grundlæggende at vurdere studerendes sandsynlighed for at besvare et spørgsmål korrekt som funktion af deres kompetencer og spørgsmålets sværhedsgrad og herefter at vurdere, om denne sandsynlighed er signifikant forskellig for mænd og for kvinder. Der ses ingen klar tendens i disse resultater. I 5 spørgsmål ses ingen forskel, i 12 spørgsmål en forskel i

mændenes favør og i 13 spørgsmål en forskel i kvindernes favør. Af disse 25 spørgsmål er det kun 3 af dem, i hvilken der ses det forfatterne kalder en "substantial DIF", nemlig spørgsmålene 14 og 23 der favoriserer mænd og spørgsmål 15 der favoriserer kvinder. Interessant nok er det netop i spørgsmål 14 og 23 der ses den største kønsforskelse i (Dockett & Heller, 2008). (Popp et al., 2011) undersøger resultaterne af FCI hvis disse 3 spørgsmål fjernes fra scoringen. Dette ændrer ikke på, at der ses en kønsforskelse i de gennemsnitlige FCI-scoringer, hvorfor det konkluderes at der på baggrund af dette studium ikke er grund til at tro, at kønsforskellen skyldes en indbygget bias i FCI som måleinstrument.

Vi ser altså opsummerende, at den veletablerede kønsforskelse i FCI-resultater ikke er særlig velforklaret hvilket påpeger behovet for videre undersøgelser.

## 7 Analyse af FCI-data fra NBI

Dette kapitel har til formål at besvare specialets første forskningsspørgsmål, nemlig

- At undersøge, hvorvidt den i udlandet observerede kønsforskel genfindes på NBI og at undersøge hvilken rolle de studerendes udeladte svar har i forhold til FCI-resultaterne og at undersøge, om der ses en kønsforskel her der kan være med til at forklare den observerede kønsforskel i FCI.

Dette gøres ved først at redegøre for de metodiske overvejelser der ligger til grund for analysen. Herefter analyseres først om der eksisterer en kønsforskel i de samlede FCI-scorer, herefter, hvordan kønsforskellen fordeler sig på de enkelte FCI-spørgsmål og slutteligt, hvor stor en indflydelse blanke svar har på FCI-resultaterne.

### 7.1 Kvantitativ metode

Dette kapitel beskæftiger sig med analyse af kvantitativ data fra Niels Bohr Institutet fra årene 2008 til 2011.

For hver årgang består data af de enkelte studerendes svar på alle 30 spørgsmål i FCI-testen. Jeg har valgt kun at kigge på data fra prætesten som er gennemført lige omkring studiestart. Herudfra er der udregnet den samlede procentuelle score for hver enkelt studerende. Disse rå data for hver enkelt årgang er vedlagt i bilag.

I følgende tabel er angivet hvor meget data der har været til rådighed.

	2008	2009	2010	2011	Antal i alt
Antal kvinder	20	35	32	49	136
Antal mænd	38	69	113	145	365
Antal i alt	58	104	145	194	501

Figur 4 Tilgængelige FCI-data fra Niels Bohr Institutet fordelt på henholdsvis mænd og kvinder og årene 2008 til 2011

Det observeres, at årgangen med start i 2008 er bemærkelsesværdig lille – faktisk har den kun lige godt halvt så mange studerende som den næstmindste årgang. Udover forskellen i antallet af optagne fra år til år skyldes forskellen i antallet af datapunkter sandsynligvis, at FCI-testen kun på de sidste to af de undersøgte årgange er foretaget i forbindelse med skemalagt undervisning og ikke i forbindelse med aktiviteter af mere frivillig grad som introduktionsforedrag eller lignende. Det giver potentielt en bias i forhold til, hvilke studerende der har taget testen de første to årgange.

I kapitel 7.4 analyseres det, hvorvidt der eksisterer en forskel på mandlige og kvindelige studerendes tendens til at udelade svar i FCI-testen. Dette gøres ved hjælp af en  $\chi^2$ -test. En  $\chi^2$ -test anvendes til at bekræfte eller forkaste en nulhypotese. Dette sker ved at opstille en tabel over data ud fra de variable, hvis sammenhæng skal undersøges – i dette tilfælde køn og forekomst af blanke svar. Herefter udregnes de forventede celletal i denne tabel under nulhypotesen og  $\chi^2$ -værdien udregnes som et mål for forskellen på de forventede og de faktiske celletal. Denne værdi sammenlignes herefter med en fraktil i  $\chi^2$ -fordelingen – hvilken fraktil afhænger af, hvilke signifikansniveau man er interesseret i (Bryman, 2012a, s. 349).

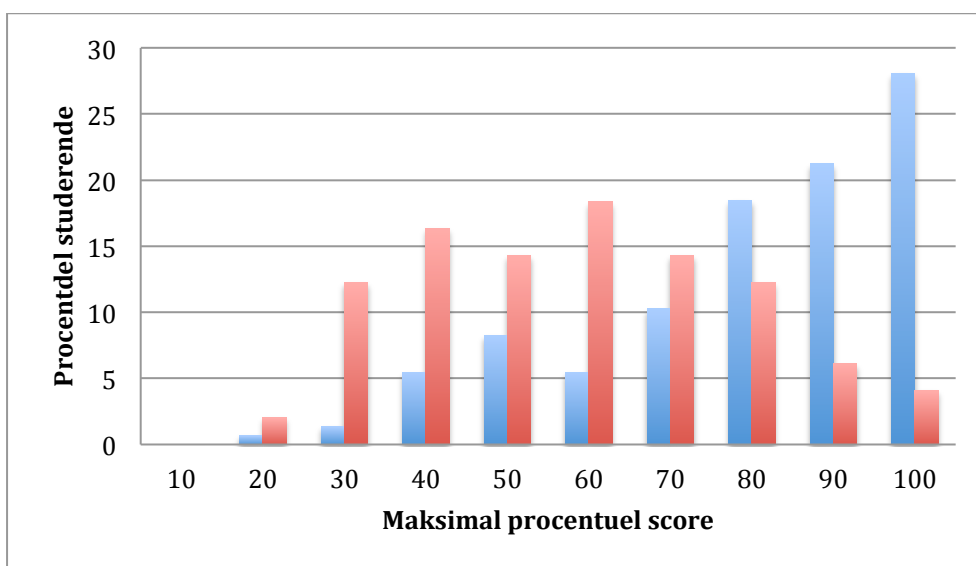


Slutteligt skal det i forhold til den kvantitative metode bemærkes, at perspektivet på, hvordan kønnet har en indflydelse på FCI-resultaterne i denne del af analysen er temmelig traditionelt, som gennemgået i kapitel 4. Dette skal dog ikke problematiseres yderligere her.

## 7.2 Analyse af scorer fordelt på køn

I det følgende ønsker jeg at analysere tidligere FCI-data fra NBI med det formål at afgøre, hvorvidt den kønsforskel, der er observeret på udenlandske uddannelsesinstitutioner genfindes her.

For hver årgang er disse data analyseret ved at gruppere henholdsvis de mandlige og de kvindelige studerende i intervaller efter deres FCI-score. Første bin går fra 0% til 10% korrekte svar, næste fra 11% til 20% og så fremdeles op til sidste bin som indeholder de studerende med 91% til 100% korrekte svar. Herefter er der plottet histogrammer over disse bins for at give et overblik over, hvordan de studerende fordeler sig. Disse er vedlagt som bilag og histogrammet for 2011 ses herunder.

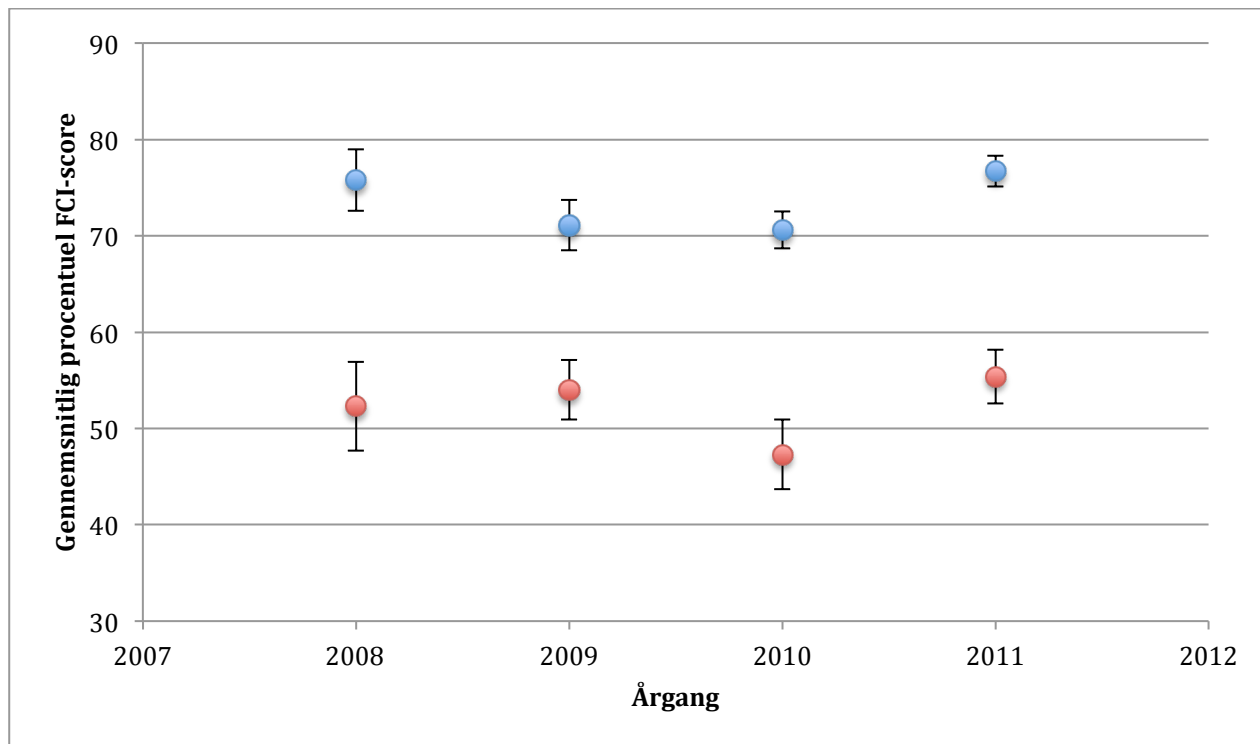


Figur 5 Procentuel fordeling af mandlige (blå) og kvindelige (røde) studerende i intervaller ud fra FCI-score. Data er fra Niels Bohr Institutet

Ved første øjekast ser disse fordelinger temmelig specielle ud. Det er udtalt, at de kvindelige studerende på hver årgang fordeler sig med markant lavere scorer end de mandlige studerende gør.

Desuden ses det, at data for hvert køn tilsyneladende er tilnærmelsesvis normalfordelt omkring to forskellige middelværdier når man tager højde for to ting. Det er ikke muligt at få over 100% korrekte. Derfor vil der ske en ophobning i den yderste højre bin hvis middelværdien ligger højt. Det er heller ikke plausibelt, at der er mange studerende der scorer mellem 0 og 20% - med de 5 forskellige svarmuligheder til hvert spørgsmål vil rent gæteri resultere i en score på omkring 20%. Derfor kan vi også tilsvarende forvente en ophobning af studerende et sted mellem 20 og 30% hvis middelværdien ligger lavt. Disse histogrammer giver altså en mistanke om, at den i udlandet observerede kønsforskel genfindes på NBI i prætestene i de undersøgte år.

For at undersøge, hvorvidt dette ikke blot er et resultat af statistiske tilfældigheder, men faktisk afspejler en signifikant forskel i kønnenes scorer, er der for hver årgang udregnet den gennemsnitlige score for henholdsvis mandlige og kvindelige studerende. Resultatet er plottet her:



Figur 6 Gennemsnitlig procentuel FCI-score for henholdsvis mænd (blå) og kvinder (røde). Error bars repræsenterer 1 standard error. Data er prætestresultater fra Niels Bohr Institutet fra årene 2008 til 2012.

Usikkerhederne på gennemsnittene er her sat til at være 1 standard error of the mean, altså standardafvigelsen divideret med kvadratroden af antallet af observationer.

Der ses altså at være markant forskel på gennemsnittene for de to køn, selv når man tager højde for, at den gennemsnitlige kvindelige score er behæftet med en vis statistisk usikkerhed på grund af de relativt få datapunkter. Denne kønsforskelse er for de 4 år som følger:

- 2008: 23,5±7,8%
- 2009: 17,1±5,7%
- 2010: 23,3±5,5%
- 2011: 21,3±4,4%

Selvom størrelsen af kønsforskellen svinger en anelse fra år til år er disse udsving ikke statistisk signifikante. Dette betyder, at vi i relation til undersøgelser, hvor vi kigger på kønsforskellen i FCI tillader os at slå data fra alle 4 år sammen.

Gøres dette fås en gennemsnitlig mandlig FCI-score for alle 4 år på 73,7±1,1% og en gennemsnitlig kvindelig score på 52,7±1,7%. Dette svarer altså til en kønsforskelse på 21,0±2,8% eller 6,3±0,8 spørgsmål.

Der ses altså en betragtelig kønsforskelse i FCI-resultater på NBI – større end både den kønsforskelse der er rapporteret i (Birch & Walet, 2012) og i (Docktor & Heller, 2008).

En mulig forklaring på denne kønsforskel kunne være, at den afspejler, at mændene har højere adgangsgivende karakterer og at kønsforskellen i FCI derfor blot skyldes en helt generel forskel i niveau blandt de to grupper af testede studerende. En sådan hypotese viser sig dog hurtigt ikke at holde stik. Således er de gennemsnitlige adgangsgivende kvotienter i 2011  $8,9 \pm 0,3$  for mændene og  $9,3 \pm 0,3$  for kvinderne. Statistisk må vi altså konkludere, at der ikke er nogen forskel på mændenes og kvinders adgangsgivende karakterer. Derudover er det værd at overveje, om en lignende kønsforskel eksisterer i eksamenskaraktererne fra Fysik 1, det kursus på NBI i klassisk mekanik og speciel relativitetsteori som FCI-testen gives i. I 2011 fik mændene gennemsnitligt karakteren  $4,4 \pm 0,3$  og kvinderne gennemsnitligt karakteren  $3,9 \pm 0,5$ .<sup>4</sup> Heller ikke her ses der nogen statistisk signifikant kønsforskel.

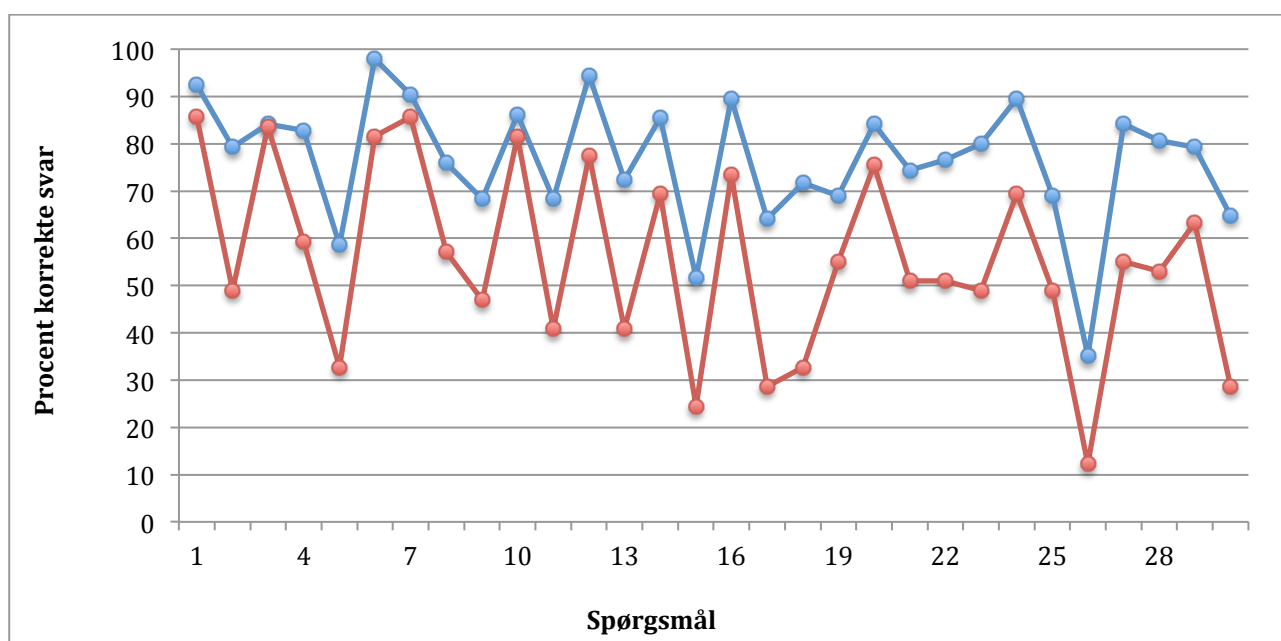
I relation til specialets første forskningsspørgsmål kan vi altså konkludere, at der eksisterer en statistisk signifikant kønsforskel i mændenes favør i FCI-data fra NBI. Disse observationer er nogenlunde konstante fra år til år. Derfor er det temmelig bemærkelsesværdigt, at denne kønsforskel hverken genfindes i adgangsgivende karakterer fra de gymnasiale uddannelser eller i de første eksamenskarakterer på fysikstudiet på NBI. Dette understreger nødvendigheden af at forsøge at forstå, hvad denne kønsforskel i FCI-resultater skyldes.

---

<sup>4</sup> Data om adgangsgivende kvotienter og eksamenskarakterer stammer fra Ian Bearden, Niels Bohr Institutet.

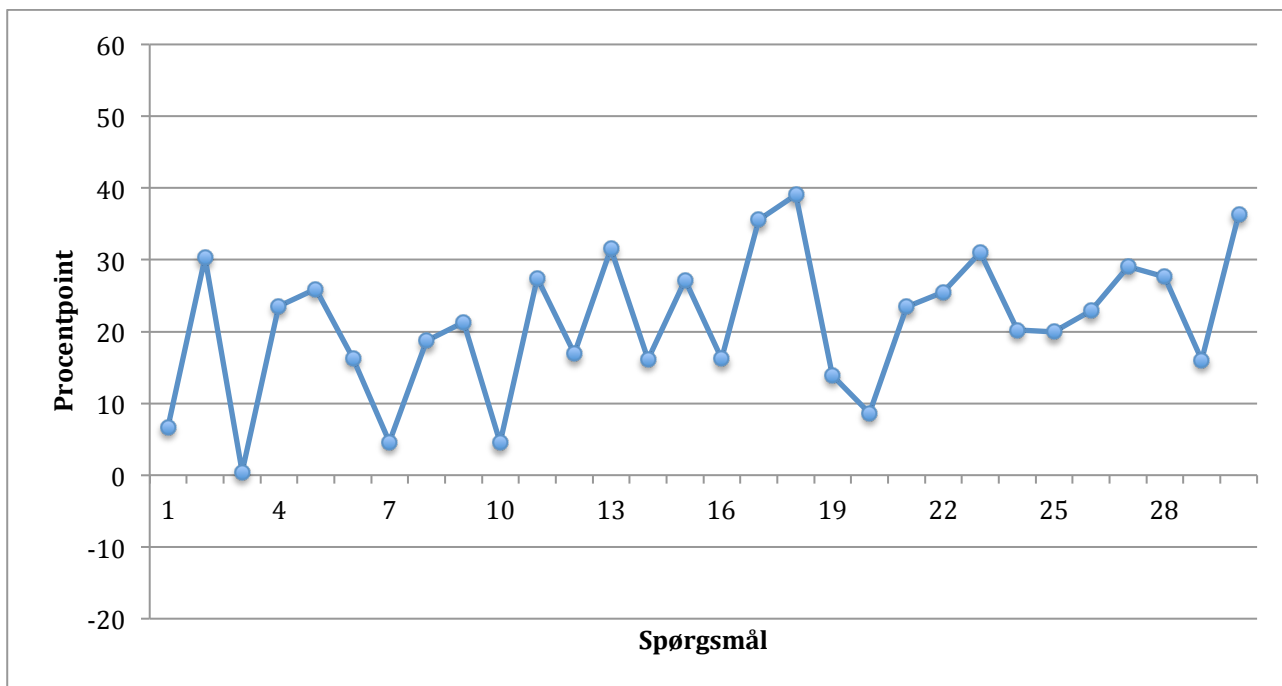
### 7.3 Analyse af kønsforskellen fordelt på de enkelte spørgsmål

I forhold til udvælgelsen af spørgsmål til brug i undersøgelsens sociokulturelt orienterede evaluering er det undersøgt, hvordan forskellen på de mandlige og kvindelige studerendes svar er fordelt. Dette er gjort ved for hvert spørgsmål at plote procentdelen af henholdsvis mandlige og kvindelige studerende der har svaret korrekt på det pågældende spørgsmål. Det er gjort for hvert af årene 2008 til 2011. På disse plots kan det yderligere aflæses, hvilke spørgsmål der overordnet set forekommer lettest eller sværest for de studerende. Herunder ses plottet for 2011. De resterende plots er vedlagt som bilag.



Figur 7 Procentuel fordeling af korrekte svar fordelt på de 30 FCI-spørgsmål for hhv. kvinder (rød) og mænd (blå). Data fra prætest på Niels Bohr Institutet fra 2011.

Helt overordnet ses altså, at kønsforskellen i mændenes favør går igen på stort set alle spørgsmål. I 2011 er der et enkelt spørgsmål, hvor andelen af korrekte svar er den samme hos de to køn, i 2010 har samtlige 30 spørgsmål en højere gennemsnitlig mandlig end kvindelig score og i 2009 og 2008 er der hvert år tre spørgsmål, hvor forskellen er i kvindernes favør. Overordnet må vi altså konkludere, at billedet her er det samme som det er observeret i Manchester og Minnesota: kønsforskellen går igen på tværs af stort set alle spørgsmål. For at lette aflæsningen af kønsforskellen på disse plots er der ydermere udarbejdet plots over forskellen i procentpoint på den mandlige og den kvindelige score. Denne forskel er udregnet som gennemsnitlig mandlig score fratrukket gennemsnitlig kvindelig score, hvorfor et positivt tal indikerer, at procentdelen af mænd der svarer korrekt er større end procentdelen af kvinder og vice versa for negative tal. Et sådant plot for 2011 ses herunder. De resterende 3 plots er vedlagt som bilag.



Figur 8 Forskel i procentpoint på andelen af mandlige hhv. kvindelige studerende der svarer korrekt fordelt på de 30 FCI-spørgsmål. Data fra 2011.

For 2011 ser vi altså, at spørgsmålene 2, 13, 17, 18, 23 og 30 udmærker sig ved at have den største kønsforskel, nemlig på 30 procentpoint eller derover. For 2008 er der tale om spørgsmålene 5, 8, 9, 13, 14, 21, 22, 23, 24 og 20, for 2009 er det spørgsmålene 5, 13, 14 og 23 og for 2010 spørgsmålene 2, 11, 13, 14, 21, 23, 25, 27 og 30.

Det ses altså, at spørgsmål 13 og 23 går igen alle 4 år, mens 14 og 30 optræder i 3 ud af de 4 analyserede år. Dette er bemærkelsesværdigt, da netop spørgsmål 14 og 23 nævnes i både (Birch & Walet, 2012) og (Docktor & Heller, 2008) som spørgsmål i hvilke kønsforskellen er størst.

Der er naturligvis statistiske fluktuationer forbundet med disse plots og der ses da også spørgsmål der kun optræder enkelte år. Nærmere statistiske usikkerhedsberegninger skal dog ikke interessere os yderligere eftersom vi ikke er interesserede i en indiskutabel eftervisning af hvilke spørgsmål der besidder den største kønsforskel. Ovenstående analyse skal snarere tages som en indikation af, hvilke spørgsmål der er ud fra et kønsperspektiv kunne være mest interessant at undersøge nærmere.

## 7.4 Analyse af udeladte svar

(Ben-Shakar & Sinai, 1991) beskæftiger sig med en hypotese om, at kvinder gennemsnitligt vil være mindre risikovillige end mænd når de testes i et multiple choice-format. Det betyder, at de hyppigere end mænd vil undlade at besvare spørgsmål frem for at komme med et (kvalificeret) gæt, hvilket alt andet lige er en uhensigtsmæssig svarstrategi i et evalueringsformat hvor man ikke straffes for forkerte svar.

I (Ben-Shakar & Sinai, 1991) ses en gennemgående tendens til flere udeladte svar hos kvinder end mænd i multiple choice-tests på tværs af testindhold og kønsforskelle i præstationen i testen. Til gengæld ses det også, at *"only a very small fraction of the observed gender differences in test performance can be accounted for by differences in guessing tendencies."*

I det følgende undersøges det, om dette billede genfindes i FCI-data.

Inspireret heraf er de studerendes blanke svar i FCI i årene 2008-2011 analyseret for at få et billede af om kønsforskellen eller dele heraf kan forklares ved nogle kvindelige studerendes uhensigtsmæssige svarstrategier med en overhyppighed af blanke svar i forhold til gennemsnittet. Data fra disse år er analyseret samlet for at skaffe en større grad af statistisk signifikans. Argumentationen for meningsfuldt at kunne slå data sammen er den samme som i kapitel 7.2 om den grundlæggende analyse af tidligere års kvantitative data.

Blanke svar i FCI kan repræsentere to ting: Den studerende kan bevidst have undladt at besvare spørgsmål som han ikke var sikker på svaret på eller han kan være løbet tør for tid og derfor ikke have nået at besvare alle spørgsmål.

Det er naturligvis ikke muligt med usvigelig sikkerhed ud fra svarene i FCI at ræsonnere sig frem til bevæggrunden for de blanke svar, så den følgende analyse tager udgangspunkt i følgende grundantagelse, der ikke er ufejlbarlig, men som overordnet burde give et billede af hvad der er på færde.

Studerende, der har udeladte svar i FCI, inddeles i to grupper. Hvis alle de blanke svar ligger i en ubrudt linje i testens slutning, således at for eksempel 4 blanke svar er placeret fra spørgsmål 27 til spørgsmål 30, antages det, at der er udeladt svar som følge af tidsmangel. Dette kaldes blanke svar af type I. Hvis der derimod er et eller flere udeladte svar andre steder i testen end til slut betragtes det som en indikation af en svarstrategi hvor den studerende udelader at svare på spørgsmål frem for at gætte. Dette kaldes blanke svar af type II.

Analysen er således foretaget ved først at afgøre, hvilke studerende der har udeladt svar og herefter placere dem i en af de to ovenfor beskrevne grupper. Herefter er det opgjort, hvor mange henholdsvis kvindelige og mandlige studerende der er i hver gruppe.

Disse resultater er som følger:

	Type I-svar	Type II-svar	Ingen blanke svar	I alt
Kvinder	22	15	99	136
Mænd	51	14	300	365
I alt	73	29	399	501

Figur 9 Faktiske celletal

For at vurdere, om det er samme fordeling der ligger til grund for henholdsvis kvinder og mænds fordeling i disse kategorier foretager vi en  $\chi^2$ -test af nulhypotesen om, at der ikke er forskel på mænds og kvinders tendens til at udelade svar i FCI-testen. De forventede celletal under denne hypotese udregnes. Resultatet ses herunder:

	Type I-svar	Type II-svar	Ingen blanke svar
Kvinder	19,8	7,9	108,3
Mænd	53,2	21,1	290,7

Figur 10 Forventede celletal under nulhypotesen

Efterfølgende udregnes  $\chi^2$ -teststørrelsen. Denne har værdien 10,2 og er større end 99%-fraktilen i en  $\chi^2$ -fordeling med 2 frihedsgrader. Vi kan altså afvise nulhypotesen med et signifikansniveau på 0,01.

Konklusionen bliver derfor, at det statistisk må antages, at der er forskel på mandlige og kvindelige studerendes tendens til at udelade svar i FCI-testen, idet kvinder er mere tilbøjelige til at udelade svar. Mest udpræget ses det i forhold til udeladelser af type II, hvor afvigelserne fra de forventede celletal er størst. Spørgsmålet er nu, om denne tendens kan forklare kønsforskellen i FCI.

I den ovenstående analyse er der heller ikke taget hensyn til, hvor mange blanke svar der udelades af hver enkelt studerende. Hvis mange kvindelige studerende udelader få spørgsmål, mens færre mandlige studerende til gengæld udelader flere spørgsmål hver vil det ikke nødvendigvis resultere i en lavere score for kvinderne. Det undersøges derfor, om der er forholdsmæssigt flere blanke svar hos et af kønnene end hos det andet.

Når alt data slås sammen har kvinderne i alt udeladt 218 spørgsmål. Da der er 136 kvinder, betyder det, at hver kvinde gennemsnitligt har udeladt 1,6 spørgsmål. De 365 mænd har derimod tilsammen udeladt 407 spørgsmål, svarende til gennemsnitligt 1,1 spørgsmål pr. mand. Udregnes på samme måde den statistiske usikkerhed på disse resultater som 1 standard error får vi at kvinderne gennemsnitligt har undladt at besvare  $0,5 \pm 0,5$  spørgsmål mere end mændene. Denne forskel forklarer altså højst en temmelig beskedne del af den observerede forskel på  $6,3 \pm 0,8$  spørgsmål.

Denne analyse fortæller intet om årsagerne til udeladelserne af svar. Muligvis udelader kvinderne oftere svar end mændene som udtryk for en forsigtighed og uvilje mod at gætte sig frem, hvilket giver dem en lavere score end de egentlig burde kunne opnå. Omvendt kan det tænkes, at de blanke svar netop er et udtryk for, at kvinderne har svært ved spørgsmålene og at deres score derfor ikke ville forbedres selvom de blev tvunget til at undlade at svare blankt. Det kunne muligvis analyseres ved at se på, om der er forskel på, hvor stor en andel af de ikke-korrekte svar de blanke svar udgør for de forskellige studerende ud fra det rationale, at jo højere denne andel er, desto større er sandsynligheden for, at de blanke svar ville blive besvaret korrekt hvis de var blevet besvaret. Dette er dog grundlæggende ikke særlig interessant ud fra et perspektiv der handler om at forstå kønsforskellen i FCI-data. Resultaterne af analysen viser netop, at højst en brøkdel af denne forskel kan forklares ud fra hypotesen om flere blanke svar hos kvinderne, der altså må forkastes som en generel forklaringsmodel.

Til gengæld afspejler analysen en forskel i kønnenes tilgang til evalueringsformatet – der er flere kvinder end mænd der har uhensigtsmæssige svarstrategier. Dette peger altså i retning af, at det også ud fra et kønsperspektiv kan være interessant at undersøge, hvad en ændring af evalueringsformatet betyder for FCI-resultaterne.

Et helt andet resultat af denne analyse er, at den sætter fokus på, at nogle studerende har en betragtelig mængde blanke svar i FCI – der findes på alle de undersøgte årgange eksempler på studerende med mere end 10 ud af 30 blanke svar og hvor disse blanke svar udgør over

halvdelen af alle de ikke-korrekte svar. Fortolkningen af disse studerendes FCI-scoringer må tages med et kraftigt gran salt når der for eksempel findes en studerende med en score på 47%, hvilket dækker over 14 rigtige, 1 forkert og 15 blanke svar. Ved for uforbeholdent at sætte sin lid til FCI-scoren som mål for mekanikkompetence risikerer man at overse, at for eksempel læsekompetencer og evne til at strukturere sin tid tilsyneladende kan influere ganske betragteligt på resultatet.

## 7.5 Konklusion

Vi har altså set, at der på alle de fire undersøgte årgange fra Niels Bohr Institutet eksisterer en statistisk signifikant kønsforskel i FCI-resultater som mildest talt er overraskende, da en sådan kønsforskel ikke eksisterer i adgangsgivende karakter og der heller ikke er en statistisk signifikant kønsforskel i eksamenskarakter i det kursus, hvor FCI-testen gives ved kursusstart. Da en lignende kønsforskel er observeret i udlandet er det væsentligt at forstå, hvad den skyldes. Det ses ligeledes ud fra ovenstående analyse, at kønsforskellen ikke kan forklares som en overhyppighed af blanke svar hos de kvindelige studerende. Til gengæld ses det, at flere kvinder end mænd har uhensigtsmæssige svarstrategier i den forstand at de udelader svar i stedet for at gætte.

Disse konklusioner gør, at det er værd at overveje, om kønsforskellen i FCI-resultaterne kan ændres ved at ændre på evalueringsformatet. At gennemføre en undersøgelse, der tager udgangspunkt i en sådan ændring bliver derfor formålet med resten af denne specialrapport i relation til indledningens andet forskningsspørgsmål.



## 8 Gentestning af FCI-spørgsmål – empiriindsamling og metode

Efter således at have konkluderet, at der også på Niels Bohr Institutet ses en kønsforskel i FCI-resultater er formålet med det følgende afsnit at beskrive de metodiske overvejelser der ligger til grund for den empiriindsamling der skal bruges til at besvare specialets andet forskningsspørgsmål. Dette er som bekendt følgende:

- At undersøge, hvad der sker med resultaterne af FCI-testen hvis evalueringsformen ændres fra en traditionel multiple choice-test til en sociokulturelt orienteret evaluering og at forstå, hvad denne resultatændring skyldes.

Denne undersøgelse er interessant i egen ret som en evaluerings- og læringsteoretisk undersøgelse af FCI-testen, men den relaterer sig også til specialets første forskningsspørgsmål og konklusionerne på dette. For det første indikerer overhyppigheden af kvinder med blanke svar i FCI-testen at der kan være en kønsforskel i tilgangen til et bestemt evalueringsformat. Desuden er en af konklusionerne på VAP-projektet (Dolin & Krogh, 2011), at en gentestning af PISA-spørgsmål i et sociokulturelt orienteret evalueringsformat forårsager den største resultatforbedring for den fagligt svageste elevgruppe, der i FCI-sammenhæng har en overrepræsentation af kvinder. Også i forhold til at blive klogere på kønsforskellen i FCI kunne det altså være interessant at genteste et udvalg af fysikstuderende ved at stille dem FCI-spørgsmål i et nyt evalueringsformat.

En beskrivelse og diskussion af de metodiske overvejelser der er gjort under arbejdet med dette speciale er temmelig omfattende. Dette overordnede metodeafsnit er derfor inddelt i flere underafsnit.

En vigtig inspirationskilde under udarbejdelsen af specialets metode har været VAP-projektet. Jeg vil derfor starte med kort at introducere dette projekt, således at det kan danne en baggrund som mine metodiske overvejelser kan holdes op mod.

Hernæst beskrives case studies som metodisk udgangspunkt ligesom udvælgelsen af cases begrundes.

Herefter gennemgås selve udarbejdelsen af det nye evalueringsformat, hvilket indebærer en diskussion af udvælgelsen af relevante spørgsmål, af, hvordan sociokulturelle aspekter er inkorporeret i gentestningen og af, hvordan interviewene er struktureret.

Herudover beskrives, hvordan interviewene helt konkret blev afholdt, ligesom metodiske vanskeligheder omkring sammenligningen mellem to grundlæggende forskellige evalueringsparadigmer diskuteres.

### 8.1 Validering af PISA-projekt

Dette speciale har fundet betydelig inspiration i Validering af PISA Science-projektet (forkortet VAP, (Dolin & Krogh, 2011)) både rent metodisk, men også i høj grad i forhold til i det hele tage at få ideen til at undersøge, hvad der sker med de studerendes præstation i FCI-spørgsmål ved en ændring af det evalueringsparadigme som testningen udfolder sig under. Det er derfor nødvendigt med et overblik over, hvad formålet med VAP-projektet er, hvilke konklusioner man kan drage fra det og hvordan det metodisk er opbygget.

PISA er en forkortelse for Programme for International Student Assessment og er OECD's testning af 15-årige inden for områderne læsning, matematik og naturfag. PISA's første testrunde fandt sted i 2000 og er siden gennemført hvert tredje år. Det helt overordnede formål med VAP-projektet er ifølge (Dolin & Krogh, 2011, s. 1) at placere PISA 2006 science-testen "i forhold til de danske mål for naturfagsundervisningen". Det gøres via tre forskningsspørgsmål, der afdækkes i hver sin rapport. Jeg har i forhold til dette speciale primært interesseret mig for det tredje af disse spørgsmål, nemlig:

*"Er PISA-resultaterne et validt udtryk for, hvad danske elever faktisk kan indenfor det testede område?"* og den dertilhørende VAP-rapport nummer 3, (Dolin & Krogh, 2011). Dette forskningsspørgsmål udfoldes herefter til to uddybende spørgsmål, hvoraf det første lyder: *"Hvor meget vil PISA-resultatet ændres, såfremt elevernes formåen hvad angår originale PISA-opgaver efterprøves indenfor et mere sociokulturelt evalueringsparadigme med dialog, adgang til medierende artefakter og mulighed for konkret praksis?"* (Dolin & Krogh, 2011, s. 34), idet PISA-testen i sin oprindelige form opfattes som tilhørende et traditionelt postpositivistisk evalueringsparadigme. Det andet forskningsspørgsmål for dette speciale ses altså at lægge i direkte forlængelse af dette forskningsspørgsmål for VAP-projektet.

Der er naturligvis en verden til forskel på omfanget af VAP-projektet, hvor der er indsamlet data nok til at opnå statistisk signifikans i en kvantitativ databehandling, og omfanget af dette specialeprojekt. Alligevel har jeg under udarbejdelsen af den metode, som min empiriindsamling baserer sig på, søgt inspiration i VAP-projektet, da nogle af de mere generelle, metodiske overvejelser stadig er relevante.

I VAP-projektet er et udvalg af folkeskoleelever blevet testet i dels en sociokulturelt funderet samtale om PISA-opgaver og dels en praktisk opgave. Det er samtalen, der er det interessante her. Denne samtale tager udgangspunkt i et samtaleskema baseret på en faglig analyse af opgaveindholdet i relation til PISA's scoringskriterier og inddrager forskellige medierende artefakter undervejs.

Konklusionen på VAP-projektet bliver, at i en direkte sammenligning efter PISA's scoringskriterier klarer eleverne sig cirka 25% bedre i det sociokulturelt orienterede evalueringsparadigme end de gør i PISA-testen. Desuden konkluderes det, at *"VAPs sociokulturelle setup er således klart mere gunstigt/inkluderende for de svagere elever"* (Dolin & Krogh, 2011, s. 57).

Disse konklusioner peger altså på, at en ændring af evalueringsformatet også kunne tænkes at have en indflydelse på resultaterne af FCI-testen, der jo, jævnfør kapitel 5.6.2, ligeledes er hjemmehørende i et postpositivistisk evalueringsparadigme.

## 8.2 Case studies

Et grundlæggende metodisk valg i forhold til dataindsamlingen handler om antallet af studerende der skal interviewes. Grundlæggende kan det ses som en afvejning mellem to yderpunkter: mange studerende, hvilket giver adgang til en stor mængde data og dermed større sandsynlighed for statistisk at kunne konkludere noget ud fra de indsamlede data, men som til gengæld er tidskrævende at gennemføre og færre studerende, hvilket næppe giver statistisk signifikans, men som omvendt giver mulighed for en mere dybdegående analyse af hvert enkelt interview.

At søge at opnå statistisk signifikans i forhold til at afgøre, hvilken betydning en ændring af evalueringsparadigmet har på FCI-scoren er omfattende og helt uden for rammerne af dette projekt. Jeg besluttede derfor relativt tidligt i forløbet at gøre en dyd af nødvendigheden og

anlægge et andet metodologisk udgangspunkt for undersøgelserne, således at projektet giver mulighed for en mere dybdegående analyse af færre udvalgte cases.

Således forskydes fokus i analysen af den indsamlede empiri også fra en kvantitativ til en kvalitativ analyse. Dette er dog ikke ensbetydende med, at det ikke ud fra case study-undersøgelserne kan være muligt at drage konklusioner angående evalueringsformatets betydning for FCI-præstationen. (Flyvbjerg, 2006) argumenterer for, at der godt kan konkluderes ud fra enkeltstående cases såfremt de udvælges ud fra bestemte kriterier der afhænger af, hvad der skal undersøges. Udgangspunktet for udvælgelsen af cases bliver ifølge (Flyvbjerg, 2006) enten en tilfældig udvælgelse eller en informationsorienteret udvælgelse. I en tilfældig udvælgelse udvælges informanterne tilfældigt, hvilket kan være hensigtsmæssigt hvis man gerne vil opnå viden om et gennemsnitligt eller typisk medlem af populationen. I en informationsorienteret udvælgelse er målet at maksimere mængden af information man får om hele populationen ud fra de undersøgte cases. De cases der knytter sig til den informationsorienterede udvælgelse inddeler (Flyvbjerg, 2006) i 4 typer: maksimum variation, ekstreme, kritiske og paradigmatisk cases. I maksimum variation-cases søger man at opnå information om betydningen af, at en given parameter ændres mest muligt. I relation til dette speciale kunne en sådan parameter for eksempel være FCI-score. Ekstreme cases er cases, der på den ene eller anden måde konstituerer noget usædvanligt, og som derfor giver en information om bredden af det undersøgte område, og formålet med paradigmatisk cases er ifølge (Flyvbjerg, 2006) *"to develop a metaphor or establish a school for the domain that the case concerns"*. Slutteligt er kritiske cases cases, der muliggør logiske slutninger; som for eksempel, at hvis et givet kriterium gælder for den betragtede case gælder det for alle andre cases

I forhold til denne undersøgelse er særligt en type af cases interessant: maksimum variation hvad FCI-score angår muliggør en undersøgelse af, om der er faktorer der er forskellige hos informanter med henholdsvis høj og lav score. Dette kan være interessant i forhold til at overveje hvilke faktorer, der kan have en indvirkning på FCI-scoren.

I forhold til at overveje ændringen af evalueringsparadigmets indflydelse på FCI-præstationen kunne det være hensigtsmæssigt at afgøre, hvilke cases der kunne betragtes som kritiske. Da der er mange faktorer der kan have indflydelse på, hvordan man præsterer i de forskellige evalueringsparadigmer er det vanskeligt at opstille kriterier for, hvad der kunne udgøre kritiske cases. I VAP-projektet ses der evidens for, at de svageste elever drager størst fordel af en ændring af evalueringsformatet. Dette giver os en formodning om, at hvis vi ser en ændring for de studerende, der har den højeste FCI-score vil den også ses for studerende med en lavere oprindelig score. I forhold til at opstille kritiske cases kunne det derfor være oplagt vælge cases blandt gruppen med maksimal score. Her kommer mætningsproblematikken dog til at spille uhensigtsmæssigt ind. Man vil næppe opleve nogen nævneværdig ændring hos denne gruppe af studerende – ikke på grund af evalueringsformatet, men ganske simpelt fordi, at deres oprindelige FCI-scoring er for høje til at kunne forbedres. Dermed bliver det ikke muligt at udvælge kritiske cases på denne måde. Strategien i forhold til udvælgelsen af cases bliver derfor i stedet at søge maksimum variation i forhold til FCI-score blandt de udvalgte cases således at der udvælges repræsentanter med scoringer der dækker hele spektret. Ligeledes blev informanterne udvalgt efter et ønske om maksimum variation i forhold til køn. Der blev derfor udvalgt lige mange mandlige og kvindelige cases.

Efter disse mere teoretiske overvejelser omkring case-udvælgelse er det dog værd at nævne, at der også er et mere pragmatisk element med i overvejelserne, nemlig hvilke studerende der i det hele taget var villige til at deltage i forsøget. Særligt i forhold til de kvindelige studerende

har det i en vis grad været et spørgsmål om at undersøge de cases, der rent faktisk var til rådighed.

### 8.2.1 Case studies og køn

Der har altså været nogle jordnære overvejelser angående for eksempel specialets tidsperspektiv der har haft indflydelse på det metodiske fundament. Muligheden for med de kvalitative undersøgelser at undersøge enkelte cases mere dybdegående for at forstå evalueringsparadigmets betydning er dog en klar fordel ved dette metodevalg, der ikke nødvendigvis havde været mulig i en kvantitativ tilgang til problemfeltet.

Valget af den kvalitative metode har dog også en helt anden fordel der relaterer til specialets kønsfokus. I kvantitative studier omhandlende køn bliver netop informanternes køn typisk det kriterium de inddeles efter. Disse studier er derfor, som beskrevet i afsnit 4, typisk ikke sensitive over for forskelle internt i kønnene. Derimod giver case study-metodikken mulighed for at betragte atypiske studerende i form af studerende hvis FCI-score ligger langt fra deres køns gennemsnitlige score. Således indgår der i dette studium både en kvindelig studerende med en maksimal score og en mandlig studerende hvis score er sammenlignelig med de laveste scorere hos kvinderne i undersøgelsen. Dermed åbnes der for en mere nuanceret forståelse af kønnets betydning for FCI-præstationen i stedet for at betragte alle mænd henholdsvis kvinder som samlede grupper defineret af deres køn.

### 8.3 Udvalgelsen af cases

For at udvælge cases til brug i dataindsamlingen startede jeg med at skaffe mig adgang til et sample af studerende, som kunne være interesseret i at deltage i forsøget. Testinterviews<sup>5</sup>, som blev gennemført umiddelbart inden studieårets start afslørede en tendens til bias der skulle modvirkes, nemlig, at studerende er mere tilbøjelige til at ville deltage i et interview der handler om at evaluere deres fysikfaglige viden hvis de forventer at kunne klare det godt. De færreste opsøger frivilligt en evaluering de ikke præsterer godt i. Da samplet af studerende skulle tilvejebringes skulle forsøget altså præsenteres på en måde, der ikke afskrækkede de fagligt svagere studerende fra at melde sig, men som omvendt heller ikke blev decideret misvisende for interviewenes indhold. Jeg besøgte derfor en forelæsning i første studieuge og holdt et kort oplæg om mit speciale, hvor jeg fortalte, at jeg kort fortalt var interesseret i studiestarten og i at undersøge, hvilke forudsætninger, de nystartede studerende havde med fra ungdomsuddannelserne. Derudover understregede jeg, at det var vigtigt for min undersøgelse med et så bredt udvalg af mulige interviewpersoner som muligt og opfordrede alle til at melde sig uanset alder, køn, type af ungdomsuddannelse og så videre. Ydermere undlod jeg bevidst at nævne undersøgelsens kønsperspektiv ud fra en betragtning af, at informanternes bevidsthed om dette ville risikere at frembringe en kønnethed i deres måde at agere på i interviewsituationen, der ellers ikke ville have været til stede.

Dette oplæg resulterede i et sample på 35 studerende, hvilket gav en vis frihed i forhold til udvalgelsen af cases, der nu kunne vælges ved en informationsorienteret udvælgelse med maksimum variation. Ud af disse 35 studerende var 8 kvinder og de resterende mænd. På grund af specialets oprindelige kønsfokus var det ønskeligt at interviewe nogenlunde lige mange af hvert køn, så jeg tog kontakt til alle 8 kvinder. Heraf var 2 ikke mulige at få

---

<sup>5</sup> Disse testinterviews blev gennemført med 3 andetårsstuderende fra Niels Bohr Institutet som blev fundet via et opslag i en Facebookgruppe for fysikstuderende. Disse interviews havde primært som formål at afprøve samtaleskemaet og give interviewerens mulighed for at afprøve sin interviewteknik.

arrangeret et interview med inden for den afsatte tidsramme. De resterende 6 kvinder indgår i undersøgelsen. På daværende tidspunkt var den første FCI-test netop blevet afholdt og de 6 kvinders scorer var som følger<sup>6</sup>:

- Anna: 100 %<sup>7</sup>
- Mette: 77%
- Nina: 53%
- Signe: 40%
- Karoline: 37%
- Sarah: 37%

Der ses altså en fordeling hvor en studerende har opnået maksimal score, en befinder sig lige under den Newtonske tænkningsgrænse (Hestenes & Halloun, 1995), en ligger tæt på tidligere års middelværdier for kvindelige studerende og de sidste tre studerende har en nogenlunde ens, temmelig lav score.

Da det ud fra testinterviewene blev observeret, at informationsværdien i interviews med studerende med en høj score var begrænset (de svarede fyldestgørende og korrekt og ændrede ikke forklaringer i forhold til deres oprindelige svar) blev det vurderet, at der kun var behov for en enkelt mandlig studerende fra denne gruppe. De mænd, der havde meldt sig til at ville deltage i undersøgelsen havde dog for en stor dels vedkommende temmelig høje scorer. De 9 mandlige medlemmer af samplet med de laveste scorer blev derfor kontaktet for også at opnå maksimum variation blandt de mandlige cases. Heraf lykkedes det at få interviews i stand med 6<sup>8</sup>. Som det fremgår var deres scorer ikke helt så lave som de deltagende kvinders – en deltager har en score der er sammenlignelig med de laveste kvindelige scorer, mens de resterende fire befinder sig i en middelgruppe. Slutteligt er Søren medtaget som en repræsentant for den højstscorerende gruppe. Scorerne var som følger:

- Søren: 80%
- Tobias: 67%
- Martin: 67%
- Lars: 63%
- Peter: 57%
- Jakob: 43%

Det er værd at bemærke, at denne udvælgelse er foretaget ud fra de studerendes samlede FCI-score. Da denne ikke nødvendigvis er repræsentativ for de studerendes score i de 6 spørgsmål der indgår i interviewet må man tage inddelingen af de studerende efter niveau med et gran

---

<sup>6</sup> Alle navne er anonymiseret.

<sup>7</sup> Jeg blev efterfølgende opmærksom på, at Anna begyndte på fysikstudiet på NBI i 2011, men droppede ud i løbet af efteråret for at starte igen i 2012. Hendes forudsætninger er derfor på nogle punkter anderledes end de andre deltageres.

<sup>8</sup> Der er altså gennemført interviews med 7 mandlige informanter. Den ene informant havde dog ikke deltaget i FCI-testen i 2012, men i stedet i 2011 hvor han blev optaget på fysikstudiet og senere droppede ud for at starte igen i 2012. Det blev efterfølgende vurderet, at det ville give for store usikkerheder i sammenligningen med de andre informanter, hvorfor dette interview blev udeladt af de videre undersøgelser.

salt. For eksempel har Mette besvaret alle 6 udvalgte spørgsmål korrekt i den oprindelige FCI og kan derfor ikke forventes at forbedre sin score i gentestningen selvom hendes oprindelige FCI-score ikke er maksimal.

Opsummerende kan det altså omkring udvælgelsen af cases siges, at det for de kvindelige studerendes vedkommende dybest set var et spørgsmål om at tage hvad man kunne få – der var ikke mange kvinder i det indledende sample at vælge imellem. For mændenes vedkommende var det vigtigt at få fat i medlemmerne af samplet med lave FCI-scorer for at opnå maksimum variation – mændene i samplet havde generelt forholdsvis høje FCI-scorer. Disse havde dog kun i beskedent grad meldt sig som forsøgsdeltagere, hvorfor de helt lavtscorende mænd ikke er helt så rigt repræsenterede i empirien som det kunne have været ønskeligt.

De udvalgte cases med scorer der fordeler sig mellem 37% og 100% korrekte svar i FCI repræsenterer derfor en høj grad af variation i forhold til FCI-scorer, med det forbehold, at der ikke optræder nogen af de studerende som har opnået de allerlaveste scorer (16 studerende har i 2012 opnået en score på 30% eller derunder).

#### 8.4 Udvalgelse af spørgsmål

I forhold til udvælgelsen af spørgsmål var der en del forskelligartede hensyn at tage. Disse er beskrevet i det følgende.

Først og fremmest var det væsentligt, at interviewsene havde en passende længde. Målet var at kunne afholde hvert interview på maksimalt 30 minutter inklusiv de afsluttende, ikke direkte fysikrelaterede spørgsmål. Jeg vurderede derfor, at et sted mellem 4 og 6 spørgsmål ville være passende, afhængig af deres omfang og forventede sværhedsgrad.

Desuden var det ønskeligt med et godt flow i interviewene. Dette blev tilgodeset ved at vælge spørgsmål der relaterede til hinanden, således at der ikke var for mange skift undervejs i interviewet.

Ydermere skulle de anvendte spørgsmål være nogle, der var velegnede til at oversætte til et sociokulturelt orienteret evalueringsparadigme, for eksempel ved at muliggøre brugen af artefakter og ved at være autentiske i forhold til en hverdagskontekst og informanternes skolehverdag som deres fysikviden er tillært i. Disse overvejelser uddybes i kapitel 8.5. Slutteligt var det ud fra metodologiske overvejelser hensigtsmæssigt med spørgsmål der udgjorde ekstreme cases i forhold til kønsforskellen. Ved at anvende spørgsmål med størst mulig kønsforskel kan man ved at observere en formindskelse af denne forskel sandsynliggøre, at den overordnede forskel vil mindskes i det nye paradigme. Dog var det ikke ønskeligt med spørgsmål, der havde en for høj andel af korrekte svar, da det kunne give en mætningsproblematik hvor en ændring ikke ville kunne observeres da størstedelen af interviewpersonerne allerede ville have besvaret spørgsmålet korrekt.

Konkret foregik denne udvælgelse ved at udvælge de spørgsmål fra analysen af tidligere data der havde den mest markante kønsforskel (over 30 procentpoint) og herefter anvende de øvrige kriterier til at vælge blandt disse spørgsmål.

Som det ses i kapitel 7.3 er de umiddelbart mest interessante spørgsmål 13, 14, 23 og 30. Spørgsmål 13 og 30 relaterer begge til samme type misconception, nemlig impetusforståelsen. 23 er relateret til andre misconceptions og indgår i en gruppe af 4 sammenhængende spørgsmål, nemlig spørgsmålene 21, 22, 23 og 24. Ved at vælge disse 6 spørgsmål, blev interviewets fysikfaglige del naturligt inddelt i to dele. Dermed indeholdt interviewet spørgsmål der dækkede flere typer af misconceptions, som sikrede et godt flow

uden for mange voldsomme skift og overgange og som for de flestes vedkommende havde en betydelig kønsforskel i den oprindelige FCI i tidligere år uden at det betød, at der sås en udpræget tendens til mætning for de mandlige studerende.

## 8.5 Sociokulturelle aspekter

Den metodisk ikke helt lette opgave med at udvikle et testformat, der på den ene side er sociokulturelt orienteret, men samtidig lukket nok til at være sammenlignelig med FCI's oprindelige testformat vil blive gennemgået i det følgende. Disse overvejelser relaterer sig til VAP-projektets metodiske grundlag i den forstand at der også her er tale om en oversættelse fra et evalueringsparadigme til et andet.

For til at gøre testformatet sociokulturelt orienteret skal karakteristika ved det sociokulturelle evalueringsparadigme som det er gennemgået i kapitel 3.4 inkorporeres i testformatet. Den vigtigste ændring der er foretaget handler om at gøre evalueringsformatet dialogorienteret i erkendelse af, at sproget i den sociokulturelle forståelse af læring udgør et helt centralt medierende artefakt. De studerende sættes således ikke kun i stand til at anvende sproget til at mediere deres forståelse med. De får i det dialogorienterede format også mulighed for at mediere deres forståelse via interaktion med interviewerens som repræsentant for den fysiske kulturs etablerede forståelse – et interaktivt element, der karakteriserer det sociokulturelle evalueringsparadigme og som ikke er til stede i FCI's oprindelige evalueringsformat.

Dette dialogorienterede format er også med til at præge gentestningen mod det sociokulturelle paradigme da evalueringen er situeret i en kontekst der er autentisk i den forstand at samtale og interaktion efter al sandsynlighed har spillet en væsentlig rolle i den skolehverdag, som testens fysikfaglige indhold er tillært i.

Et andet væsentligt aspekt ved oversættelsen til det sociokulturelt orienterede evalueringsparadigme handler om muligheden for at anvende materielle artefakter og altså ikke kun sproget som symbolsk artefakt undervejs i samtalen.

I spørgsmål 13 havde informanterne således en bold til rådighed, i spørgsmål 30 en bold og en ketsjer, og i spørgsmål 21 til 24 et udklippet papirrumskib, de kunne bevæge i forskellige baner. Disse artefakter er karakteriseret ved at være intuitivt tilgængelige. Der er ikke læringslag forbundet med selve artefaktet der skal gennemtrænges af den studerende – der er ingen tvivl om, hvordan man betjener en bold. Således intenderer gentestningen ikke at undersøge informanternes evne til at anvende fagets artefakter, men lader blot artefakterne være redskaber for elevernes forklaringer og betydningsdannelse undervejs i samtalerne – i modsætning til VAP-projektet, hvor evnen til at anvende for eksempel en globus udgør en vurdering af en elevs kunnen i egen ret.

Herudover giver evalueringsformatet mulighed for at afsøge elevernes nærmeste udviklingszoner, særligt udpræget i relation til spørgsmål 21, hvor en del af spørgsmålet bliver at undersøge, om informanterne er i stand til at blive klogere på spørgsmålet ved at blive cuet til at tænke i baner af et formentlig velkendt, fysisk ækvivalent eksempel.

Alle disse træk ved designet er med til at karakterisere gentestningen som orienteret mod det sociokulturelle evalueringsparadigme efter parametrene opstillet i Figur 2. Det er dog væsentligt at understrege, at der snarere er tale om en sociokulturelt orienteret evaluering end en evaluering der er fuldstændigt hjemmehørende i det sociokulturelle paradigme. Dette er et bevidst, metodisk valg, der skal legitimere den efterfølgende sammenligning af gentestningen med de oprindelige FCI-resultater.

## 8.6 Interviewenes struktur

Muligheden for sammenligning sikres ved at søge at give gentestningen tilstrækkelig høj reliabilitet i de postpositivistiske evalueringsparadigme FCI-testen er udviklet i. Dette blev imødekommet ved at lade interviewene være semistrukturerede (Bryman, 2012b). Interviewene tog, som VAP-testen (Dolin & Krogh, 2011), således udgangspunkt i et samtalskema, baseret på den fysikfaglige analyse af spørgsmålene og vedlagt i Appendix D, der har til formål at sikre, at alle relevante dele af de studerendes faglige viden blev afdækket under interviewet.

Når interviewene betegnes som semistrukturerede dækker det over, at der var en klar plan for, hvilke spørgsmål informanterne skulle stilles og interviewene er dermed tæt på at være så strukturerede, som kvalitative interviews kan blive. Interviewene kan dog ikke betegnes som kvantitative. En del af spørgsmålene er åbne eller semiåbne (som for eksempel når der i relation til spørgsmål 21 startes med spørgsmålet *"Hvordan kan det være at det (rumskibet) kan bevæge sig når der nu ikke virker nogen kræfter på det?"*) og samtalskemaets struktur er ikke fastere, end at aspekter som informanten finder særligt vanskelige, interessante eller lignende diskuteres grundigere og dermed er med til at styre samtalsforløbet.

Det blev dog besluttet, at de studerende ved slutningen af hvert spørgsmål blev bedt om at udpege det korrekte svar blandt de 5 oprindelige svarmuligheder fra FCI. Dette muliggjorde en helt direkte sammenligning mellem besvarelser i FCI-testen og i gentestningen.

Gentestningens reliabilitet højnes således, da det ikke er muligt at fortolke hvordan besvarelserne skal karaktergives, hvilket er med til at legitimere gentestningens metode i den naturvidenskabelige tradition FCI typisk anvendes i. På den måde tilføjes der et kvantitativt aspekt til de ellers kvalitative interviews. Dette afspejler de metodiske vanskeligheder ved at oversætte den postpositivistiske, kvantitative FCI-test til et kvalitativt evalueringsformat, således at en sammenligning stadig var mulig. I VAP-projektet (Dolin & Krogh, 2011, s. 49) introduceres således to forskellige scoringsmodeller – i den ene model får eleverne ligeledes mulighed for at forholde sig til PISA's svarmuligheder undervejs i interviewet og i den anden scores elevens præstation efterfølgende uden at de har haft adgang til svarene. Der argumenteres for, at den første mulighed giver eleverne *"en begunstigelse, i form af den indtuning og stilladsering som samtale-optakten leverer"*. Denne effekt søges begrænset ved at udelukke tilfælde, hvor dette har været særlig udpræget, fra resultaterne af gentestningen. En nærmere diskussion af dette ses i kapitel 9.2.

Interviewene blev indledt med et kort oplæg fra interviewerens side, hvor der blev fortalt om, hvordan interviewene skulle anvendes i relation til specialet<sup>9</sup> og om hvordan interviewet ville blive optaget og efterfølgende anonymiseret. Informanten blev desuden opfordret til at tænke højt omkring sine overvejelser omkring spørgsmålene og til at anvende artefakterne. Desuden blev informanten spurgt om navn, alder, gymnasial uddannelse og forældres uddannelsesniveau. Af anonymitetshensyn er denne del af interviewene ikke transskriberet. Efter den semistrukturerede del blev interviewene afrundet med en åben del. Formålet var primært at få et indblik i, hvorvidt informanten kunne huske opgaverne fra tidligere. Dette blev udgangspunkt for en samtale, hvis forløb primært blev struktureret af informantens svar. Der var således ikke noget samtalskema, men interviewerens havde i baghovedet at emner som blandt andet informantens opfattelse af FCI-testens sværhedsgrad, opfattelse af, hvordan intuition og hverdags erfaringer påvirker deres viden om mekanik og oplevelse af, om konkrete, autentiske opgaver er lettere, var interessante.

---

<sup>9</sup> Hverken specialets kønsfokus eller FCI-testen blev nævnt.



Denne afsluttende del af interviewene er transskriberet, men kun anvendt i ganske begrænset omfang i analysen. Dette skyldes dog snarere behovet for fokus i den begrænsede plads i specialet end manglen på interessante overvejelser hos informanterne.

### 8.7 Forskningsinterview eller evalueringssamtale?

Gentestningen blev altså planlagt som hvad man med inspiration fra VAP-projektet kunne kalde en semistruktureret sociokulturelt orienteret samtale. Denne beskrivelse dækker over vanskelighederne ved at karakterisere den indsamlede empiri fuldstændigt entydigt som evaluering – den befinder sig snarere i et krydsfelt mellem forskningsinterviews og evaluering. Der er et klart og vægtigt evalueringselement i empirien i og med at den har til hensigt at vurdere de studerendes mekanikforståelse (eller manglende forståelse) i relation til nogle bestemte opgaver, præcis som formålet med FCI er. Omvendt er der også elementer i empirien, der ikke normalt er til stede i en evalueringssituation. Således bliver informanterne i starten af interviewene bedt om at tænke højt undervejs – de får at vide, at alle deres overvejelser, også overvejelser som de ikke er sikre på er korrekte, er relevante og interessante, hvilket man ikke ville gøre i en mere sædvanlig evalueringssituation. Billedet af, hvad de studerende kan, er dermed muligvis anderledes i disse samtaler end det ville være, hvis der var tale om samtaler, der mere entydigt lod sig karakterisere som evaluering. Grundlæggende må man dog sige, at det at samtalerne ikke entydigt er evalueringer, men også har karakter af at være interviews, næppe overspiller de studerendes kundskaber, i den forstand, at man med dette udvidede opmærksomhedsvindue i højere grad får indblik i de studerendes tvivl og faglige usikkerheder end et format, der mere entydigt kan karakteriseres som evaluering, ville give.

Denne dobbelttydighed ses også i forhold til muligheden for at anvende interviewene i den senere analyse. Som evaluering betraget anvendes de transskriberede samtaler til at vurdere, om de studerende er i stand til at besvare FCI-spørgsmålene korrekt med henblik på en sammenligning med de oprindelige FCI-resultater. Betragtes samtalerne derimod i højere grad som forskningsinterviews åbner det for andre analytiske perspektiver. Adgangen til information om de studerendes verdensbillede giver mulighed for en læringsteoretisk forståelse af, hvordan dette verdensbillede konstrueres.

Under interviewene var det en målsætning, at interviewerens rolle var forholdsvis afventende således at elevernes egen forståelse kom på banen uden at blive alt for kraftigt farvet af interviewerens indblanding – igen for at legitimere samtalerne som evalueringer i en naturvidenskabelig tradition. Når det er sagt, var der dog omvendt netop tale om en sociokulturelt orienteret samtale, designet med udgangspunkt i et socialkonstruktivistisk læringssyn, hvor dialog og interaktion er væsentlige elementer, når informantens forståelse skal evalueres. Således er en fuldstændig tilbagetrukket rolle fra interviewerens side hverken mulig eller ønskelig inden for det paradigme evalueringen er udført under.

Desuden vidste interviewerens ikke under interviewene, hvad informanterne havde svaret på de enkelte spørgsmål i FCI-testen. Dette var ud fra en forestilling om, at dette kunne have en uønsket effekt i forhold til at gøre interviewerens forudindtaget i forhold til eksistensen af bestemte misconceptions hos den studerende, hvilket kunne præge interviewene i en bestemt retning.

### 8.8 Dataindsamling

Specialets empiri blev udført i starten af blok 1, altså umiddelbart efter studiestart for de studerende. Konkret fandt det første interview sted 10. september 2012, altså i 2. kursusuge og det sidste interview 18. september 2012 i kursusuge 3.

At finde et hensigtsmæssigt tidspunkt at udføre interviewsene på var et spørgsmål om at finde et fornuftigt kompromis mellem flere modsatrettede hensyn. De nystartede studerende tog FCI-testen til første laboratorieøvelse der blev afholdt i løbet af uge 36 (3.-7. september). Af hensyn til pålideligheden af denne test og dermed muligheden for at sammenligne den med testresultater fra andre år var det vigtigt, at der ikke var studerende, for hvem nogle af spørgsmålene i FCI var en gentestning af spørgsmål de havde set før. Omvendt var det vigtigt at empirien ikke blev udført efter undervisningen havde været i gang for længe, da jeg jo netop var interesseret i FCI-resultatet for de nystartede studerende og ikke ønskede at forholde mig til, hvordan universitetsundervisningen påvirker testresultatet<sup>10</sup>. Kompromiset blev derfor at starte interviewene i 2. kursusuge. Ideelt set havde jeg foretaget alle interviews i løbet af få dage, således at der ikke skete en udvikling i dels hvad der var blevet lært i universitetsundervisningen, dels hvor godt de oprindelige FCI-spørgsmål kunne genkaldes. Omvendt ville jeg gerne være så fleksibel som overhovedet muligt i forhold til de interviewede sådan at interviewene ikke forstyrrede deres øvrige undervisning eller krævede, at de mødte op på universitetet på alt for skæve tidspunkter. Dette betød, at interviewene blev strukket over godt en uge.

Interviewene blev afholdt i undervisningslokaler på H.C. Ørsted-instituttet og på et kontor og i et mødelokale på Niels Bohr-instituttet og varede mellem 15 og 30 minutter.

Jeg har selv udført interviewene. Dette havde den helt klare fordel, at jeg selv bedst vidste, hvordan jeg gerne ville have interviewene til at forløbe. Først at skulle forklare en interviewer, hvordan dataindsamlingen skulle gennemføres og sikre, at vedkommendes forståelse af, hvad der var vigtigt var konsistent med min ville give unødigt besvær og potentielt give empiri der ikke matchede hvad jeg ønskede at undersøge. Disse fordel opvejer ulemperne ved, at jeg ikke var empiriindsamlingens start var nogen særlig erfaren interviewer.

Det første interview er gennemført med Jakob. Da alle 3 testinformanter havde ganske let ved at besvare spørgsmålene er dette interview således det første med en fagligt mere usikker informant, hvilket gør, at intervieweren er noget mere usikker på interviewformatet her end i de efterfølgende interviews. Jeg har valgt alligevel at lade interviewet indgå i den videre analyse, fordi Jakob er væsentlig som repræsentant for den fagligt svageste gruppe af mandlige studerende, men det er værd at være bevidst om denne metodiske usikkerhed. Interviewene blev optaget på video, således at det også blev dokumenteret, hvordan interviewpersonerne anvendte de medbragte artefakter. Herefter blev interviewene anonymiseret og transskriberet efter principper der er beskrevet i (Bryman, 2012b).

## 8.9 Metodiske vanskeligheder

Der har været en del metodiske spidsfindigheder forbundet med udarbejdelsen af det sociokulturelt orienterede evalueringsformat såvel som med vurderingen af, hvilket analytisk blik der efterfølgende skulle lægges på empirien. Dette handler om rent faktisk at anlægge et nyt evalueringsparadigmatisk udgangspunkt, men samtidig ikke bevæge sig længere væk fra

---

<sup>10</sup> Dette er særlig relevant i forhold til kønsforskellen i FCI, da mandlige studerende gennemsnitligt har et højere gain i FCI-scoren i løbet af blok 1 end kvindelige studerende har. En for sen testning ville derfor gøre det vanskeligt at afgøre om en ændring af kønsforskellen skyldtes ændringen af evalueringsformatet eller en eventuel kønsbias i undervisningen.

FCI's lærings- og evalueringsteoretiske udgangspunkt, end at det stadig er muligt at sammenligne resultaterne fra gentestningen med de oprindelige FCI-resultater.

Dette betyder, at der er taget en del forbehold for at legitimere sammenligningen dels i det evalueringsparadigme, FCI-testen er hjemmehørende i, men også i det naturvidenskabeligt prægede metodesyn, som er en del af den videnskabelige verden, det fysiske fagfællesskab, som FCI-testen tænkes anvendt i.

Populært sagt er de metodiske overvejelser gjort med et ben i hver lejr: dels i en naturfagsdidaktisk verden, hvor det er bredt accepteret at betragte læring og evaluering i et sociokulturelt perspektiv og hvor kvalitative metoder er udbredt praksis og dels i et naturvidenskabeligt paradigme, hvor objektivitet og kvantitative metoder er idealet.

Disse to verdener er videst muligt søgt medieret under udarbejdelsen af dette speciale ud fra det rationale, at naturfagsdidaktiske undersøgelser af FCI naturligvis må tage udgangspunkt i en naturfagsdidaktisk metodik, men at konklusionerne får en større vægt ved også at have en vis legitimitet inden for fysikfaglige kredse.

Dette ses for eksempel ved, at det validitetsbegreb, der diskuteres i specialet, primært tager udgangspunkt i (Hestenes et al., 1992)'s egen forståelse af, hvad validitet vil sige uden at dette problematiseres yderligere. Ligeledes blev informanterne bedt om at vælge et svar på de oprindelige svarkort for at give mulighed for en helt direkte sammenligning af resultater i de to testformater – også selvom disse svarkort ikke nødvendigvis er særlig karakteristiske for et sociokulturelt evalueringsformat. Ligeledes tages der i analysen mange forbehold for at imødekomme kritik af, at svarændringerne skyldes, at informanterne er blevet hjulpet frem til den korrekte besvarelse. En diskussion af, hvorvidt der forekommer for meget stilladsering undervejs i evalueringen er således ikke særlig relevant set gennem de sociokulturelt prægede briller. I dette perspektiv er stilladsering en uomgængelig del af en evaluering der søger at afdække elevens nærmeste udviklingszoner. Ligeledes er svarændringerne holdt meget konservative i den forstand, at de er udeladt fra analysen ved tvivl om, om informanten er blevet stilladseret for kraftigt, om det svar der er valgt faktisk afspejler en reel forståelse hos informanten eller hvis informanten har været for tøvende i forhold til at lægge sig fast på en af flere valgmuligheder.

## 9 Resultater

Efter at have gennemført og transskriberet interviewene (vedlagt som bilag) har vi en betragtelig mængde data som de følgende kapitler tjener til at analysere.

I dette kapitel præsenteres det, hvilke besvarelser der er givet i gentestningen og det analyseres, hvordan de adskiller sig fra besvarelsenerne der er givet i det oprindelige testformat.

### 9.1 Sammenligning af svar

I interviewene blev de studerende i slutningen af hvert spørgsmål bedt om at vælge et svar blandt de oprindelige fem svarmuligheder i FCI. Dette muliggør en direkte sammenligning mellem de studerendes svar i de to testformater.

I det følgende gives der et overblik over, hvilke ændringer af svar der sker fra FCI-testen til gentestningen. Dette er gjort ved at notere informanternes svar i de to testformater i et skema, der er vedlagt som bilag. Disse resultater er konservative i den forstand, at i de tilfælde hvor informanten har svært ved at lægge sig fast på et svar eller hvor der kan stilles spørgsmålstejn ved, om det er tale om en for voldsom stilladsering eller guidning fra interviewerens side der gør, at informanten ændrer sit svar, er der angivet en usikkerhed på svaret i form af et spørgsmålstejn. Ændringen fra FCI til gentestningen er derfor ikke mulig at bestemme entydigt, hvilket er angivet i skemaet. Et eksempel kunne være Peter, der i relation til spørgsmål 13 siger følgende:

*"Så 3'eren, den kan vi også godt godkende. Skal vi lige.. Så er der 1 og 3 tilbage lige at vælge i mellem og så 4, en stort set konstant gravitationskraft. Mmm, det er nok den som mit intellekt siger, men min intuition strider imod. (...) Ja, altså, jeg må jo nok sige at jeg har lavet den samme fejl sådan cirka 20 gange, så jeg bliver nok nødt til at sige at det rigtige svar må være en stort set konstant gravitationskraft."*

Peter ender altså med at vælge det korrekte svar, men ikke uden også at angive, at han faktisk mener, at nogle af de andre svar også kunne være de korrekte. For at minimere usikkerheden i resultaterne af gentestningen er hans svar altså angivet som ubestemt, også selvom hans svar ville have talt som korrekt i FCI's forstand. Det betyder, at vi må tage de nedenstående resultater som et konservativt bud på, hvilke ændringer gentestningen forårsager. For at give et overblik over ændringerne betegnes de forskellige typer af ændringer ved forskellige bogstaver som angivet herunder.

- A= uændret korrekt
- B= uændret forkert
- C = ændring fra forkert til korrekt
- D = ændring fra blankt til korrekt
- E= ændring fra forkert til noget andet forkert

	Q13	Q30	Q21	Q22	Q23	Q24	Score i FCI
<b>Anna</b>	A	A	A	A	A	A	100%
<b>Jakob</b>	B	B	Ubestemt	A	C	A	43%
<b>Karoline</b>	Ubestemt	Ubestemt	Ubestemt	C	C	A	37%
<b>Lars</b>	A	D	C	C	A	A	63%
<b>Martin</b>	A	D	Ubestemt	A	A	A	67%
<b>Mette</b>	A	A	A	A	A	A	77%
<b>Nina</b>	B	E	C	C	A	C	53%
<b>Peter</b>	Ubestemt	D	C	A	A	A	57%
<b>Sarah</b>	C	D	C	C	C	A	37%
<b>Signe</b>	Ubestemt	E	Ubestemt	C	A	D	40%
<b>Søren</b>	C	C	A	A	A	A	80%
<b>Tobias</b>	C	A	Ubestemt	C	A	A	67%

Figur 11 Oversigt over ændringer af besvarelsene i gentestningen.

Vi ser her, at der i de 72 spørgsmål sker en ændring i 25 af tilfældene, svarende til 35%, ingen ændring i 38 tilfælde (53%) mens ændringen ikke er entydigt bestemt i de sidste 9 tilfælde (13%). Af de 25 ændrede svar ændres 18 fra et forkert svar til det korrekte, 5 fra et blankt svar til det korrekte og i 2 tilfælde fra et forkert svar til et andet forkert svar. Af de 38 uændrede svar er 35 uændret korrekte, mens kun 3 er uændret forkerte.

Helt overordnet kan vi altså konkludere, at ændringen af evalueringsformatet betyder noget for, hvilke svar de studerende giver på spørgsmålene – vi får adgang til andre aspekter af deres viden ved at ændre på testformatet. Dette er ikke nødvendigvis ensbetydende med et forbedret testresultat, men alene det, at der faktisk svares så forskelligt i de to formater er interessant.

Det er ydermere interessant at bemærke, at de eneste to informanter, der svarer uændret på alle spørgsmålene er Anna og Mette, der har besvaret alle 6 spørgsmål korrekt i FCI. Samtidig ses ikke et eneste tilfælde af, at et korrekt svar i FCI ændres til et forkert i samtaletesten. Dette peger i retning af, at de korrekte svar afspejler en forståelse, der er grundlæggende nok for den studerende til ikke at ændres på trods af ændringen af evalueringsformen. I alle ændringer der ses er der således tale om et enten blankt eller forkert svar der ændres til et korrekt – eller om et forkert svar der ændres til et andet forkert svar. Dette antyder altså, at de forkerte svar ikke er et udtryk for et rationelt og konsistent ikke-Newtonsk verdensbillede hos den studerende, siden de så relativt ofte ændres ved en ændring af testformatet.

I (Hestenes et al., 1992) angives det, at der forekommer falsk positive svar i FCI-testen, hvor ikke-Newtonsk tænkende studerende svarer korrekt, men at falsk negative svar, hvor en Newtonsk tænkende studerende svarer forkert er sjældne. Dette billede går tilsyneladende ikke igen i en test, hvor evalueringsparadigmet ændres som det er gjort i denne opgave. Her ses en del eksempler på, at studerende ændrer deres forkerte besvarelser i FCI-testen til korrekte besvarelser i gentestningen. Der ses også eksempler på studerende der er i stand til at vælge det korrekte Newtonske svar uden at kunne give en fyldestgørende Newtonsk forklaring af, hvorfor svaret er korrekt, altså eksempler på, hvad Hestenes et al. ville kalde falsk positive. Dette ses for eksempel når Peter ud fra en intuitiv forståelse vælger det korrekte svar på spørgsmål 21, men begrundet det således:

*"Altså, den vil gerne bevæge sig begge veje samtidig, kan man sige. (...) Det er lidt ligesom når man sender en bold af sted, så laver den også den der parabel. Måske det er fordi.. Altså ting bare bevæger sig i parabler, det ved jeg ikke.. Det må jeg indrømme."*<sup>11</sup>

Det ses altså, at billedet af, at en Newtonsk tænker kun sjældent vælger det forkerte FCI-svar, men at ikke-Newtonske tænkere somme tider vælger korrekte svar med ikke-Newtonske begrundelser ikke genfindes her. Dette peger på, at den læringsteoretiske baggrund for FCI som den er diskuteret i kapitel 5.6.1 er værd at undersøge nærmere, hvilket sker i kapitel 11.

Der ses altså en temmelig gennemgående ændring af besvarelser ved en ændring af det paradigme, evalueringen gennemføres under. For at afgøre, om nogle studerende drager større fordel af gentestningen end andre udregnes for hver studerende deres normaliserede gain<sup>12</sup>. De ubestemte svar i gentestningen tælles ikke med som korrekte for at få indblik i det mest konservative bud på resultatændringen. Desuden bør det bemærkes, at gainet ikke er veldefineret for Anna og Mette, der besvarer alle spørgsmål korrekt i både FCI og gentestningen.

	Antal korrekte svar af 6 mulige, FCI	Antal korrekte svar af 6 mulige, gentestning	Normaliseret gain	Oprindelig FCI-score
<b>Anna</b>	6	6	Udefineret	100%
<b>Karoline</b>	1	3	0,4	37%
<b>Mette</b>	6	6	Udefineret	77%
<b>Nina</b>	1	4	0,6	53%
<b>Sarah</b>	1	6	1	37%
<b>Signe</b>	1	3	0,4	40%
<b>Jakob</b>	2	2	0	43%
<b>Lars</b>	3	6	1	63%
<b>Martin</b>	4	5	0,5	67%
<b>Peter</b>	3	5	0,7	57%
<b>Søren</b>	4	6	1	80%
<b>Tobias</b>	3	5	0,7	67%

Figur 12 Normaliserede gains fra FCI til gentestningen for de 12 informanter

Vi ser helt overordnet, at der sker en ændring mod flere korrekte svar for alle informanter på nær Jakob (og Anna og Mette, som jo ikke har mulighed for at forbedre deres score).

Ændringen af evalueringsformatet har altså en betydning for størstedelen af informanterne og det bliver derfor væsentligt at søge at forstå, hvad den skyldes og hvad det betyder for vores opfattelse af FCI-testen.

At komme med nærmere statistiske overvejelser omkring ændringerne repræsenteret ved de udregnede gains er ikke meningsfuldt med den begrænsede datamængde der er til rådighed.

<sup>11</sup> Det er værd at bemærke, at dette eksempel er med til at understrege, hvorfor en skarp inddeling af studerendes tænkning som enten Newtonsk eller ikke-Newtonsk er unuanceret. Peters tankegang er netop ikke ikke-Newtonsk. Der er snarere tale om en intuitivt baseret og ikke formaliseret udfoldet Newtonsk tankegang.

<sup>12</sup> Udregnet som  $\frac{\text{postscore} - \text{præscore}}{6 - \text{præscore}}$ , hvor præscoren er antallet af de 6 spørgsmål der er besvaret korrekt i FCI og postscoren er antal spørgsmål besvaret korrekt i gentestningen.

Nogle overvejelser omkring betydningen af dels køn og dels oprindelig FCI-score er dog alligevel på sin plads.

Det ses, at der sker markante ændringer for både de medvirkende mænd og kvinder. Der er altså ikke umiddelbart noget der tyder på, at ændringen i meget udpræget grad har en større betydning for det ene køn end for det andet. Den begrænsede datamængde giver os dog ikke mulighed for at bekræfte dette via en statistisk analyse. Resultatændringen er dog så karakteristisk, at det interessante ved disse resultater i lige så høj grad bliver at forstå hvordan resultaterne skal forstås læringsteoretisk og hvilken betydning det har for vores opfattelse af FCI som at overveje, om ændringen af testformatet påvirker kønsforskellen. I kapitel 11 diskuteres årsagerne til resultatændringerne evalueringsteoretisk. Der skal vi vende tilbage til, om nogen af disse årsager kan karakteriseres som kønnede.

Opsummerende kan vi dog sige, at disse foreløbige resultater snarere gør det relevant at lade den videre analyse beskæftige sig med empirien fra et lærings- og evalueringsteoretisk perspektiv end ud fra et kønsperspektiv.

I forhold til spørgsmålet om den oprindelige FCI-scores betydning for gainet er billedet ikke entydigt. Således har Jakob, Karoline og Sarah sammenlignelige oprindelige FCI-scoringer, men vidt forskellige gains på henholdsvis 0, 0,4 og 1. Ligeledes ser vi et maksimalt gain på 1 for både Sarah, med en oprindelig score på 37% og for Søren med en oprindelig score på 80%. Den væsentligste konklusion at drage er dog, at siden vi ser maksimale gains for cases der ligger yderligt i forhold til hele populationen kan vi forvente, at ændringen af evalueringsformat potentielt kan ændre besvarelserne for alle typer af studerende og ikke kun for en gruppe af studerende med bestemte FCI-resultater.

Som tidligere nævnt fastholdes der kun i ganske få tilfælde et forkert svar i gentestningen. Dette understøtter billedet af, at forkerte svar i FCI ikke skal ses som et udtryk for et konsistent ikke-Newtonsk verdensbillede, men snarere som et ikke-konsistent Newtonsk verdensbillede, der hviler på et så ustabil grundlag at det er forholdsvis let rokkeligt. De svar, som var ubesvarede i FCI, ligger på nær et i spørgsmål 24, alle i spørgsmål 30, hvor hele 5 af interviewpersonerne oprindeligt har svaret blankt. Af disse i alt 6 blanke svar bliver 5 besvaret korrekt efterfølgende. Dette må tages som et udtryk for, at de blanke svar i FCI, der når scoren udregnes har samme status som forkerte svar, ikke nødvendigvis skal opfattes som et udtryk for manglende forståelse hos den studerende, men måske snarere som et udtryk for tidsmangel.

Det er karakteristisk, at de ubestemte svar i gentestningen primært fordeler sig på de spørgsmål 13 og 21, der tilsyneladende volder de studerende de største vanskeligheder. Der ses desuden en meget begrænset ændring for spørgsmål 24, hvilke beror på, at 10 ud af de 12 informanter har besvaret dette spørgsmål korrekt i FCI-testen.

## 9.2 Nuancering af scorer

Den kvalitative metode, der er anvendt i empiriindsamlingen giver mulighed for at opnå en dybere indsigt i, hvilke ræsonnementer og forståelser der ligger til grund for de studerendes valg af svar i FCI-spørgsmålene. Dette muliggør en undersøgelse af nystartede fysikstuderendes viden om og kompetencer inden for Newtonsk mekanik som kan nuanceres mere og udfoldes bredere end FCI-scoringerne giver mulighed for.

Denne udfoldelse kan ske på flere måder, både anekdotisk baseret og gennem en mere systematisk analyse.

Den eksplorative tilgang kan beskrives som en udforskning af udsagn eller træk ved studerendes besvarelser der er særligt overraskende eller interessante uden at der nødvendigvis er tale om træk der er gennemgående hos alle informanter. Således kunne man for eksempel gribe fat i hvordan nogle studerende blander dynamiske og kinematiske grundbegreber sammen, som når Peter bliver bedt om at forklare, hvilke kræfter der virker på bolden i spørgsmål 13 og roder grundigt rundt i kræfter, hastigheder og energi: *"Der virker øhm tyngdekraften (...) og så er der den øh.. den energi som jeg har påført den i starten, den kraft jeg har påført som man kan kalde starthastigheden, altså v nul."*

En sådan fremdragning af eksempler, der giver en indsigt i de studerendes forståelse og konceptuelle vanskeligheder ses dels i den evalueringsteoretiske analyse af gentestningen og dels mere systematisk i den læringsteoretiske analyse af spørgsmål 21 – om end der ganske givet er flere interessante aspekter der kunne gribes fat i i en eventuel senere analyse af den indsamlede empiri.

En mere systematisk analyse kunne tage udgangspunkt i en pointgivning ud fra et pointgivningsskema udviklet i en tæt symbiose med samtaleskemaet – hvad der bliver spurgt om influerer hvordan der kan gives point og hvad man ønsker at pointgive influerer hvad man spørger om. Et sådant pointskema muliggør at informanten kan tildeles forskellige pointtal alt efter for eksempel graden af stilladsering der er påkrævet for at eleven svarer fyldestgørende og informantens taksonomiske niveau – kan informanten for eksempel udelukkende gengive Newtons 2. lov eller formår han også at anvende den på den relevante situation. Når dette ikke er gjort i dette speciale skyldes det to ting:

For det første anlægger en sådan nuancering af scorerne som et mål for de studerendes forståelse et metodesyn der tager et statistisk udgangspunkt: Formålet bliver en kvantificering af de studerendes forståelse, der efterfølgende skal behandles statistisk med det mål at opnå et bredt og dækkende indblik i hele årgangens kunnen. En sådan analyse er derfor først for alvor meningsfuld når der er undersøgt tilstrækkelig mange informanter til at konklusionerne har statistisk signifikans, hvilket ikke er tilfældet med den empiri der er indsamlet i dette speciale.

For det andet er formålet med en sådan indsamling grundlæggende anderledes end formålet med dette speciale som det kommer til udtryk i de opstillede forskningsspørgsmål. Her er formålet med empiriindsamlingen at søge at forstå FCI-testen ud fra både et køns- og et evalueringsperspektiv. Formålet med en nuanceret pointgivning bliver derimod at søge at beskrive fysikstuderendes viden om mekanik, hvilket jo grundlæggende er et ganske andet formål. For at fastholde specialets fokus er en sådan pointgivning derfor udeladt.

Når dette er sagt giver det dog stadig et væsentligt perspektiv på den indsamlede empiri at muligheden for en sådan nuanceret scoring af interviewene eksisterer. Dette ses for eksempel i relation til de forholdsvis mange ubestemte besvarelser der ses i den direkte sammenligning af FCI-resultaterne med gentestningen. Når Peter således i relation til spørgsmål 13 angiver en splittelse mellem at vælge det korrekte svar, som hans intellekt fortæller ham er rigtigt, og et forkert svar, som han intuitivt drages af, er en direkte scoring af resultaterne meget følsom overfor informantens valg af svarstrategi. Peter siger således om sin svarstrategi: *"Jeg synes ikke at der er nogen grund til at svare noget man ikke forstår. Altså bare at svare som en dum maskine, det er ikke min stil. Så vil jeg hellere svare forkert indtil jeg forstår det rigtigt."* Denne splittelse mellem intellektuel og intuitiv forståelse indfanges til gengæld i en nuanceret scoring af resultaterne, der åbner for en forståelse af, at Peter befinder sig på en slags midterstadium, hvor han har tilegnet sig nogle Newtonske tankegange ved at have set



tidligere eksempler – han kan genkalde sig dem, men de er ikke meningsfulde for ham. Nuanceringen af scorerne åbner således for en forståelse af, at Newtonsk tankegang i relation til hvert enkelt spørgsmål ikke nødvendigvis er noget man enten har eller ikke har, men derimod er noget, man kan have tilegnet sig i forskellige grader.

### 9.3 Konklusion

Opsummerende kan det altså siges, at empirien viser, at det resultat vi får i de 6 udvalgte FCI-spørgsmål er afhængig af det evalueringsformat vi vælger. Denne afhængighed ses at være meget grundlæggende – der ses ændringer fordelt på hele spektret af udvalgte cases. Det videre mål bliver altså at analysere disse resultater og diskutere deres betydning for forståelsen af FCI-testen og forståelsen af, hvordan den Newtonske mekanik skal forstås i et læringsteoretisk perspektiv.

## 10 Evalueringsteoretisk analyse af den udførte empiri

Grundlæggende ses der i gentestningen mange eksempler på, at studerendes besvarelser ændres i forhold til de svar som blev givet i FCI. I det følgende diskuteres det, hvad denne ændring kan skyldes – både ud fra de teoretiske overvejelser som ligger til grund for ændringen af evalueringsparadigmet, men også ud fra nogle mere pragmatiske overvejelser. For at forstå, hvorfor ændringen finder sted ud fra et evalueringsteoretisk perspektiv må man søge at forstå, hvordan informanterne forholder sig i det ændrede evalueringsparadigme med de muligheder for anvendelse af artefakter og dialog som der er til stede her og som er diskuteret i kapitel 3.4.

### 10.1 Anvendelse af de materielle artefakter

En central pointe i det ændrede evalueringsparadigme er, at de studerende her har adgang til konkrete artefakter, der kan understøtte deres forståelse og formidling af den fysik, de testes i. De studerende kan potentielt set bruge artefakterne til at udtrykke viden igennem, som ellers ikke havde været tilgængelig for dem.

I gentestningen anvendes flere forskellige artefakter: en bold, en ketsjer, et udklippet fotografi af et rumskib og papir og blyant til at tegne med. En væsentlig pointe i forhold til disse artefakter er, at der næppe er tale om decideret læringsunderstøttende artefakter. I (Dolin & Krogh, 2011) optræder der således for eksempel en globus, som eleverne direkte kan anvende i deres videnskonstruktion, for eksempel ved at blive gjort opmærksom på jordaksens hældning. Artefakterne i FCI-gentestningen er snarere intuitivt tilgængelige og er ikke snævert knyttet sammen med formelle læringssituationer i fysik.

Der er stor forskel på hvordan artefakterne anvendes af informanterne. Både bolden og rumskibet benyttes således af alle informanter, hvorimod ketsjeren stort set ikke benyttes. Peter tager fat i ketsjeren da han forklarer om spørgsmål 30 og bruger den desuden til at illudere en kanon spørgsmål 21 – ingen af de andre informanter benytter den.

At rumskibet benyttes af alle informanterne må siges at være temmelig naturligt eftersom de eksplicit bliver opfordret til at fortsætte den bevægelse, som interviewereren har bevæget rumskibet i. Således bruges rumskibet i udpræget grad til at demonstrere løsninger på et problem med, men for nogle informanterne er der også knyttet en forståelse til denne mulighed for at bevæge rumskibet. Dette ses for eksempel hos Nina, når hun overvejer rumskibets bane ved at bevæge det rundt på papiret mens hun siger følgende:

*"Den vil jo stadig have en retning den vej (langs AB). Så skal jeg lige selv.. Det er her jeg bliver visuel (Bevæger rumskibet rundt). Og så starter den... Altså, det vil være denne her retning (peger skråt til højre), men om det bliver en buet bane eller om det bliver en skrå, det er sådan lidt (...)"*

Bolden bruges i udpræget grad til at demonstrere med, for eksempel ved at bevæge bolden mens det forklares, hvordan hastigheden ændres undervejs. I forhold til at afgøre, at der ikke virker en kraft fra hånden eller tennisketsjeren på bolden i relation til spørgsmål 13 og 30 kunne man forvente, at dette var lettere for informanterne når de helt konkret kan se bolden. Dette er dog ikke noget der nævnes eksplicit.

## 10.2 Anvendelse af sproget som artefakt

Helt centralt i den socialkonstruktivistiske læringsteori står forståelsen af, at sproget er et medierende artefakt – populært sagt lærer man i denne forståelse fysik ved at lære at tale om det. Dermed giver gentestningen de studerende mulighed for selv at konstruere mening undervejs i interviewet gennem anvendelsen af dette artefakt.

Peter er et udpræget eksempel på dette. Da han for eksempel bliver spurgt om, hvordan rumskibet i spørgsmål 21 kan bevæge sig uden der virker kræfter på det, forsøger han at konstruere mening ved at formulere sig mundtligt omkring situationen – anvendelsen af sproget som artefakt er en essentiel og uadskillelig del af videnskonstruktionen som det ses i følgende, temmelig lange, citat.

*Peter: Den skal lige gennemskues. Skal man lige lave noget blabla igen? Okay, så Newtons 1. lov siger, at summen af kræfterne er lig med 0 hvis den bevæger sig med en konstant fart. Men så er der jo stadigvæk nogle kræfter. Der kan man jo for eksempel have en øh.. noget friktion. Det har vi så ikke ude i rummet. Og noget der skubber den ene vej. Okay, den bevæger sig. Og der er ikke nogen kræfter der påvirker den. Altså, ude i rummet så forestiller jeg mig bare at der er sådan noget gravitation. Altså planeter mellem hinanden der påvirker hinanden.*

*Interviewer: Nej, men vi er så langt væk fra alting, så ikke en gang gravitations.. Ja, ingen kræfter.*

*Peter: Nårh, det er sådan noget vakuum. Den bevæger sig, og der er ikke nogen kraft der påvirker den.*

*Interviewer: Og der er ikke nogen kræfter der påvirker den.*

*Peter: Det kan jo faktisk være at vi lidt er ovre i samme boldgade som før, hvor bolden kan bevæge sig frem uden der er nogen kraft der skubber den frem. Fordi der var denne der gravitation der trak den nedad og friktionen der trak den tilbage. Så hvis nu man siger at den er blevet skudt af sted med en tennisketsjer, jamen så har du igen kun.. så har den.. der er ikke nogen kræfter der påvirker den nu.. Uh, nu tror jeg, at jeg er ved at forstå det hele. Det vil sige, at den er blevet sendt af sted og så har den en vis mængde energi i sig, for eksempel, eller den bevæger sig i en retning. Og hvis ikke der er noget der stopper den.. Det er fordi alt er i bevægelse hele tiden, men der skal være en kraft til at stoppe den. Det er derfor den kan bevæge sig frem. Okay, denne her bliver skubbet frem fordi den har en eller anden bevægelses eller hvad kan man kalde det, kinetisk energi, bevægelsesenergi og hvis ikke der er noget der stopper den, så bevæger den sig bare.*

At Peters forklaring ikke er korrekt ændrer ikke på pointen her: Hans konstruktion af mening og forståelse er bundet til at foregå via sproget.

Udover denne overordnede brug af sproget som artefakt ses der forskellige interessante træk i informanternes måde at anvende sproget til videnskonstruktion og –formidling.

Anvendelsen af sproget som et meningskonstruerende artefakt giver informanterne mulighed for at udtrykke tvivl omkring deres forståelse og ikke endegyldigt at lægge sig fast på en forklaringsmodel. Dette betyder, at vi i gentestningen muligvis får et bredere indblik i deres forståelse når de også udtrykker de dele af den, hvor de ikke er helt overbeviste om korrektheden af deres udsagn. Muligheden for at moderere sine udsagn med udtryk for usikkerhed gør muligvis informanterne mindre tilbageholdende med at formulere dele af deres viden som de er mere usikre på. Dette ses for eksempel når Tobias bevæger rumskibet og siger om bevægelsen at det er "Altså sådan, på en eller anden måde-agtigt" og dermed får mulighed for at lade denne viden være et første udspil i forståelsen af rumskibets bane – eller

hos Nina, der om boldens bevægelse i spørgsmål 13 siger *"Det er her at jeg altid får det forkert fordi jeg synes det burde gå hurtigere og hurtigere mens den kommer nedad men det er der nogle gange hvor det ikke gør"* og dermed ved hjælp af sprogets mulighed for at udtrykke sin tvivl præsenterer sin splittelse mellem sin egen forståelse og hvad hun erfaringsmæssigt ved er den korrekte forklaring. Dermed giver sproget mulighed for ideafprøvelse og mediering af informantens egen forståelse med den officielle forståelse.

Der er nogle af informanterne, der sprogligt iscenesætter de abstrakte objekter der optræder i opgaverne som handlende objekter der besidder en form for vilje. Disse antropomorferinger giver en mulighed for narrativt at forstå og forklare objekternes bevægelser. Et eksempel på dette ses hos Sarah, der skal forklare, hvordan bolden i spørgsmål 13 bevæger sig opad i starten af kastet selvom der virker en nedadrettet tyngdekraft på den. Dette gør hun således: *"(...) bolden bliver ved med at gå opad fordi den sådan lidt "nåh, okay, og så blev jeg sluppet og så..." Den bliver ved med at gå opad på grund af den hastighed som den har til at starte med."* Hun forstår altså boldens bevægelse via en form for indre vilje eller mulighed for handling hos bolden (den bliver ved med at bevæge sig opad indtil den så at sige opdager at den er blevet sluppet og reagerer på det). Denne forklaring oversættes herefter til en formaliseret, fysikfagligt acceptabel og korrekt besvarelse.

Disse antropomorferinger kan bruges til at skabe forståelse med, men informanterne er samtidig i høj grad bevidste om, at det ikke er det ikke er fysikfagligt acceptabelt ræsonnementer. Således bringer flere af dem svarmulighed E i spørgsmål 13 på bane som noget de med det samme kan udelukke. For eksempel siger Lars da han læser dette svarkort: *"5'eren den er lidt lol<sup>13</sup>, men 4'eren det er i hvert fald den rigtige,"* og da han senere bliver spurgt om, hvad han tænker om svarmulighed E siger han at *"det er jo ikke fordi at bolden i sig selv har den vilje at den vil ned mod jorden, det er jo fordi der er en tyngdekraft"*.

En væsentlig pointe er, at meningskonstruktionen gennem sproget som artefakt er afhængigt af, at man faktisk er fortrolig med brugen af dette artefakt. Således er Signe et eksempel på en studerende, for hvem kravet om anvendelsen af sproget som artefakt tilsyneladende spænder ben for videnskonstruktionen. Signe har taget en international studentereksamen og er således ikke vant til at formulere sig fysikfagligt på dansk. For hende er evalueringen desitueret i forhold til den kontekst hendes viden er konstrueret i og hun virker tøvende og uvant med brugen af fagsproget på dansk som for eksempel i følgende citat hvor hun forklarer hvilke kræfter der virker på bolden i spørgsmål 13:

*Signe: Altså, hvis ikke vi har luftmodstand så, øhm... Næh, jeg ville ikke tro der var flere end den, hvad hedder det, force som du, hvad hedder det?*

*Interviewer: Kaster?*

*Signe: Virker på... Ja, exact, on it, og så tyngdekraften ned. Det, ja...*

Sproget er tydeligvis en barriere hun skal gennemtrænge, hvilket muligvis hindrer hendes fysikfaglige viden i at komme til udtryk. I hvert fald har hun næppe meget glæde af sproget og dialogen som meningsskabende artefakt.

---

<sup>13</sup> Populær forkortelse af laughing out loud, dvs. angivelse af, at svarmuligheden er skør eller fjollet.

### 10.3 Viden som socialt situeret

En væsentlig forskel på de to evalueringsformater er, at den sociokulturelt funderede samtale i gentestningen bliver mere autentisk i den forstand, at den søger at spejle den sociale kontekst, det som de studerendes viden er konstrueret i. Ud fra et socialkonstruktivistisk synspunkt er det lærte uadskilleligt fra den proces det er lært i. Derfor vil en evalueringsform, hvor den studerendes viden testes med muligheden for at spejle forståelsen i det praksisfællesskab der udgøres af fysikfaget give et andet indblik i den studerendes kompetencer og det er altså med dette udgangspunkt ikke overraskende, at de studerende svarer anderledes i en evalueringsform, der er tættere på de læringsituationer deres viden er konstrueret i.

Det mere sociokulturelt orienterede syn på evaluering, der lader evalueringsformen spejle læringsprocessen, åbner også for en forståelse af læring og evaluering som to processer, der er ikke er mulige at adskille fuldstændigt. Netop fordi det lærte og læringsprocessen er knyttet så tæt sammen kan man ikke meningsfuldt betragte det lærte som en størrelse, der eksisterer uafhængigt af kontekst hos eleven og som derfor kan måles uden at måleprocessen influerer på det lærte. Meningsdannelsen hos eleven vedbliver med at foregå under evalueringen og læring og evaluering bliver dermed også uadskillelige størrelser. Det ses flere steder i gentestningen.

Eksempelvis oplever Peter gennem interviewet en betydelig splittelse mellem hvad han selv kalder intuition og intellekt i forhold til spørgsmål 13 og 30, hvilket for eksempel ses når han om boldens bevægelse i spørgsmål 13 siger:

*"Det er i hvert fald en kraft, der er en kraft der påvirker den. Men så er det lige som om når man laver de der opgaver, så ser man lidt sådan bort fra den."* Han er klar over, at der noget uldent ved hans impetusinspirerede forståelse af boldens bevægelse, men er ikke i stand til at give en tilfredsstillende alternativ forklaring. Denne splittelse generer ham tydeligt under interviewet og han befinder sig derfor i en læringsproces i den forstand at han vedvarende søger at forene de to modsatrettede forståelser af boldens bevægelse – det han selv kalder sin intuition og sit intellekt. Gentestningen indfanger i højere grad end FCI denne splittelse og billedet af Peters viden bliver her, at han befinder sig i en læringsproces, hvor han gradvist begynder at tænke Newtonsk i kontekster, hvor han ellers ikke har gjort det. I FCI er billedet derimod afhængigt af hvilken svarstrategi han valgte – at svare efter den intuitive eller den intellektuelle forståelse.

Måske mere udtalt ses dette hos Nina, der i højere grad end Peter formår at gennemføre læringsprocessen undervejs i evalueringen i forhold til spørgsmål 21, hvor hun ender med at være i stand til at ræsonnere sig frem til en fyldestgørende forklaring af rumskibets bane og efterfølgende konstaterer *"Ah, det er første gang jeg har forstået det."*

Det nye evalueringsformats mulighed for at lade læring foregå simultant med evalueringen tillader derfor et bredere opmærksomhedsvindue i forhold til de studerendes kunnen inden for den Newtonske mekanik.

### 10.4 Viden som kontekstuel situeret

Viden kan ikke kun være socialt situeret, men også mere generelt kontekstuel situeret. Det kan således være svært at overføre viden fra en fysisk situation eller et eksempel til andre lignende situationer. Dette ses meget udpræget i interviewene i relation til spørgsmål 21, hvor det er markant sværere for mange af informanterne at anvende de samme fysiske lovmæssigheder i relation til rumskibet end til kanonen. Dette vil blive grundigere analyseret i kapitel 11 og vil derfor ikke blive nærmere udfoldet her.

Et andet udtryk for, at viden er kontekstuel situeret ses i forhold til informanternes splittelse mellem hverdagsforståelser og mere formaliserede forståelser. De svar, der gives er afhængige af, i hvor høj grad informanterne er i stand til at oversætte mellem hverdagsforståelserne og de formaliserede forståelser. Dette ses mest udpræget i relation til spørgsmål 21, hvor der er mange mere tekniske overvejelser om rumskibet på spil. Eksempelvis svarer Tobias i relation til et spørgsmål om rumskibets hastighed når det påvirkes af en konstant kraft, at *"du ville nok bare sådan power den op til en eller anden fart og så ville den bare flyve med konstant hastighed."* Når det understreges, at motoren ikke slukkes, men bliver ved med at yde en konstant kraft, svarer han derimod, at *"Så vil du vel køre med konstant acceleration, såfremt at du holder den samme kraft."* Dette indikerer, at hans forståelsesproblemer ikke i høj så grad handler om en manglende evne til at anvende Newtons 2. lov som en skepsis i forhold til at acceptere, at et rumskib ikke stopper sin acceleration når det når sin ønskede hastighed. Dette influerer hans valg af bevægelseskurve idet han siger, at *"Hvis du kører med konstant acceleration så er jeg sikker på at det vil være denne der (E'eren), fordi så stiger du langsomt den vej i starten og så begynder det mere og mere."*

Samme overvejelser ses hos Karoline, der om rumskibets fart efter motoren er blevet tændt siger:

*"Ja, det er øhm.. nummer 2, med mindre.. altså kontinuert voksende, med mindre at øhm... Ja, men måske.. Når jeg nu tænker på hvordan det virker når man kører i bil så skal man jo først lige accelerere og så har man.. på et eller andet tidspunkt finder man så, jamen den hastighed, den er god og nu føles det godt og motoren gør nogen dejlig lyd og så bliver man jo ved med.. konstant acceleration og så ville det jo være.. Hvis rumskibet gør det samme, så ville det jo være voksende et stykke og konstant derefter, men umiddelbart at sådan det du har sagt, at motoren virker hele tiden, så er det kontinuert voksende."*

I interviewene ses også flere tilfælde af mere tekniske overvejelser omkring motoren og hvorvidt den faktisk vil yde den konstante kraft med det samme den tændes eller om den har brug for tid til at varme op. For eksempel siger Nina, at hun i spørgsmål 21 vil vælge den korrekte bane E, *"med mindre at den var meget langsom om at starte, men så tæller det vel ikke"* ligesom Jakob siger, at *"Altså, det kan ikke være den der (peger på C'eren), for den skal have tid til at reagere også, vel"*. Der er ikke spurgt nok ind til dette i interviewene til at kunne udelukke, at der er tale om en forståelse af, at alle objekter har en indbygget modvilje mod at ændre deres bevægelse (som i misconception I4 i Table II i Appendix B) men eksemplerne antyder, at der også kan være tale om nogle mere tekniske overvejelser hos informanterne der vanskeliggør deres forståelse.

Der ses altså eksempler på, at forkerte svar i FCI-testen kan bunde i hverdagsforståelser. I gentestningens evalueringsformat er det muligt at afklare, om disse forkerte svar bunder i konceptuelle misforståelser eller er et udtryk for en manglende evne til at manøvrere i en idealiseret fysikverden hvor for eksempel motorer tænder instantant eller hvor rumskibsførere ikke slukker for motoren når deres rumskib har opnået den ønskede hastighed.

### 10.5 Tids- og læsefærdighedsproblematikker

En noget mere pragmatisk forskel på de to evalueringsformater handler om den tid, de studerende har til rådighed. Hele FCI-testen gennemføres på NBI på 30 minutter og de studerende har altså gennemsnitligt 1 minut til at besvare hvert spørgsmål, hvilket giver dem markant mindre tid til at gennemtænke deres ræsonnementer og lægge sig fast på et svar end

de får i det nye evalueringsformat, hvor hvert interview (dog inklusive den afsluttende snak som ikke er en del af selve gentestningen) har taget mellem 15 og 30 minutter. Den samme problematik afspejles i de mange blanke svar som blev diskuteret i afsnit 7.4. Da alle studerende har samme tid til rådighed i FCI-testen og der ydermere ikke er nogen forskel på præ- og posttesten ændrer denne problematik ikke umiddelbart på muligheden for at sammenligne FCI-scoring på tværs af studerende eller at betragte gainet fra præ- til posttesten som et mål for undervisningens effekt på ændringen af den konceptuelle forståelse hos de studerende. Til gengæld har den en indflydelse i forhold til at anvende FCI som et absolut mål af studerendes konceptuelle forståelse i den forstand at besvarelser, der rummer mange blanke eller ikke ordentligt gennemtænkte svar ikke indfanger hvad de studerende kan præstere under mindre pressede former.

Lige så pragmatisk, men ikke desto mindre relevant er, at svage læsere stilles stærkere i gentestningen end de gør i FCI's oprindelige testformat – både fordi den begrænsede tid i det oprindelige testformat kan være en hindring for langsomme læsere, men også fordi, gentestningen giver mulighed for at sikre, at de studerende faktisk forstår de forskellige svarmuligheders fysiske indhold. Dette ses i interviewene i udpræget grad i forhold til det teksttunge spørgsmål 13. Peter gennemgår de forskellige svarmuligheder under interviewet og siger at *"(...) 3'eren, den kan vi også godt godkende. Skal vi lige.. Så er der 1 og 3 tilbage lige at vælge i mellem"*. Han er ikke ved gennemlæsningen i stand til at adskille de to svarmuligheder fra hinanden. Endnu tydeligere bliver det hos Jakob, der slet ikke formår at adskille svarmulighederne og efter gennemlæsningen blot konstaterer *"Øhm, jeg synes de alle sammen er rigtige."* Først ved interviewerens forklaring af forskelle og ligheder på de forskellige svarmuligheder formår han at genkende svarmulighed 3 som det han selv har forklaret omkring kræfterne på bolden inden han blev præsenteret for svarkortet.

### 10.6 Grupper af sammenhængende spørgsmål

Spørgsmålene 21-24 udgør en samlet gruppe, der alle forholder sig til den samme fysiske situation. Dette betyder, at et forkert besvaret spørgsmål kan have en indvirkning på, hvordan der svares i de efterfølgende spørgsmål uden at dette nødvendigvis bunder i fejlopfattelser omkring disse. Dette sker ikke i gentestningen, da informanterne får at vide, hvad det korrekte svar er. Således sikres, at for eksempel spørgsmål 23 reelt er fortæller om, hvad informanten har af forståelse om banen efter motoren er slukket og ikke fortæller om et overleveret forkert svar fra spørgsmål 21.

Et tydeligt eksempel på dette er Karoline. I gentestningen vakler hun meget mellem svarmulighed B og E i spørgsmål 21. Efterfølgende får hun at vide, at E er det korrekte svar. Hun svarer efterfølgende korrekt på spørgsmål 23 og argumenterer også for, at rumskibet ikke kan ændre retning når det ikke påvirkes af kræfter. I FCI har hun derimod valgt det forkerte svar B i spørgsmål 21 og ligeledes det forkerte svar C i spørgsmål 23. Interessant er det dog, at denne svarmulighed faktisk ville være den korrekte såfremt rumskibet faktisk bevægede sig i en bane som den hun valgte i spørgsmål 21. Således fortæller Karolines forkerte svar i spørgsmål 23 noget om hendes misconceptions i relation til spørgsmål 21 – men ikke noget om hendes forståelse af situationen i spørgsmål 23.

### 10.7 Stilladsering

I forhold til vurderingen af gentestningens validitet som mål for de studerendes kendskab til og evne til anvendelse af den Newtonske mekanik er det væsentligt at diskutere den stilladsering der finder sted under interviewene. En kritik af det evalueringsformat der

anvendes i gentestningen kunne være, at der ikke er tale om en direkte aflæsning af de studerendes kunnen, men at interviewene snarere er undervisningsforløb hvor interviewerens intervention i form af stilladsering og ledende spørgsmål ændrer informanternes forståelse undervejs i forløbet.

Denne kritik afspejler det metodisk vanskelige i at sammenligne de to meget forskelligartede evalueringsparadigmer i henholdsvis FCI og gentestningen, netop fordi blikket på stilladsering som en del af evalueringsprocessen er væsensforskellig i de to paradigmer. Idealet i det traditionelle evalueringsparadigme er netop en objektiv måling af viden som en størrelse, eleven har erhvervet sig i større eller mindre grad – hvorimod denne ide om eksistensen af et objektivt måleinstrument, der holder læring og evaluering adskilt, er meningsløs i et sociokulturelt orienteret evalueringsparadigme. En grundlæggende præmis i forhold til at sammenligne resultaterne fra de to evalueringsparadigmer bliver altså, at en vis grad af stilladsering må anses som uomgængelig del af det sociokulturelle evalueringsparadigme, snarere end som en metodefejl i interviewene.

Når det er sagt har det stadig været et metodisk mål under udførelsen af interviewene at stilladseringen blev minimeret og at interviewerens rolle var at spørge ind til informantens forståelse indtil denne var fyldestgørende afdækket, snarere end at søge at udvikle informanternes forståelse undervejs i interviewet – også selvom dette er paradoksalt i den forstand, at en evalueringsproces i en sociokulturel forståelse skal afsøge informantens nærmeste udviklingszoner og lede efter best performance og ikke typical performance. Der ses fra tid til anden eksempler på for kraftig stilladsering, ledende spørgsmål eller indblanding fra interviewerens side, for eksempel i interviewet med Jakob, hvor følgende ordveksling om luftmodstanden finder sted:

*Interviewer: Ja, eller hvordan det kan være at hvis jeg er herinde at der ikke er nogen luftmodstand på bolden?*

*Jakob: Fordi der ikke er noget luft? Nej, det ved jeg ikke. Jeg forstår ikke spørgsmålet. (Griner)*

*Interviewer: Øhm*

*Jakob: Herinde er der kun påvirkelse af tyngdekraften jo*

*Interviewer: Okay. Og udenfor, har det så noget at gøre med at det blæser mere udenfor end det gør herinde, altså at der er mere vind?*

*Jakob: Ja, der er mere vind.*

*Interviewer: Okay. Så når du siger, at der ikke er noget luft herinde så mener du måske i virkeligheden mere at det ikke blæser herinde?*

*Jakob: Ja, det blæser ikke herinde.*

I sådanne tilfælde er en ændring af besvarelsen fra det ene testformat til det andet udeladt af analysen, og selvom den momentvise uhensigtsmæssigt kraftige stilladsering fra interviewerens side kan ses som et metodisk kritikpunkt ændrer det ikke ved det grundlæggende billede af, at besvarelsene ændres fra FCI til gentestningen. Søren er således det betydeligste eksempel på en studerende, hvis besvarelse af spørgsmål 13 og 30 er så umiddelbart korrekt i gentestningen at ændringen af svarene ikke kan tilskrives nogen form for stilladsering eller indblanding fra interviewerens side. Han svarer umiddelbart korrekt, fyldestgørende og udtømmende på disse spørgsmål, selvom de begge er besvaret forkert i FCI med en angivelse af, at bolden påvirkes af en kraft fra henholdsvis hånd og ketsjer. Et særligt interessant område i forhold til diskussionen omkring stilladsering relaterer sig til spørgsmål 21. Her introduceres et fysisk set ækvivalent, men kontekstuel anderledes



eksempel til de informanter, der ikke umiddelbart besvarer spørgsmålet korrekt. Efterfølgende spørges de, om dette eksempel ændrer deres forståelse af den oprindelige situation med rumskibet, hvis bane skal bestemmes. Introduktionen af dette eksempel kan ses som en form for stilladsering, idet informanternes opmærksomhed henledes på særlige træk ved den situation der diskuteres. En nærmere læringsteoretisk analyse af dette ses i kapitel 11. Her skal det blot nævnes, at svaret betragtes som ændret i analysen hvis to krav er opfyldt: informanten ændrer sit svar til det korrekte udelukkende på baggrund af interviewerens opfordring til at gentænke situationen set i lyset af eksemplet med kanonen på bakken og altså uden yderligere stilladsering og informanten er ydermere i stand til at give en tilnærmelsesvis fyldestgørende fysisk forklaring af banen. Det er altså ikke nok, at informanten ændrer med en forklaring om, at de to baner ligner hinanden. Det betyder, at Lars er den eneste, der i analysen er angivet som at have ændret svar på spørgsmål 21 efter at være blevet bedt om først at svare på spørgsmålet om kanonkuglens bevægelse.

### 10.8 Gentestningsproblematikken

En andet aspekt omkring gentestningen i det sociokulturelt orienterede evalueringsparadigme der kan give anledning til en metodekritik er, at der netop er tale om en gentestning. Dermed kunne det tænkes, at de studerendes besvarelser ændres til det bedre, netop fordi der er tale om spørgsmål som de har set før og som de derfor oplever som mere genkendelige og dermed lettere.

Dette spørgsmål lader sig ikke entydigt besvare. Det kunne have været interessant at give FCI-testen i det oprindelige evalueringsformat til en kontrolgruppe af studerende samtidig med at gentestningen af informanterne fandt sted. Dette kunne give et indblik i, hvor stor en del af ændringen af besvarelserne der skyldes, at interviewene er en gentestning af allerede stillede spørgsmål. En sådan gentestningen ligger dog uden for rammerne af dette projekt, hvorfor vi må forholde os lidt mere løseligt til denne problematik.

Alle informanter blev i slutningen af interviewet spurgt om de kunne huske at have set spørgsmålene før, hvilket alle angav at de kunne. Nogle bragte det endda uopfordret på bane under interviewet. Det er karakteristisk, at det primært er spørgsmålene om rumraketten de studerende angiver at have set tidligere, hvilket muligvis kan hænge sammen med, at spørgsmål 21 opleves som vanskeligst og derfor i højere grad har fæstnet sig i informanternes hukommelse. Selvom alle informanterne således kan huske FCI-testen er det ikke alle spørgsmålene de kan huske herfra.

Nogle af informanterne angiver, at de har tænkt over spørgsmålene siden FCI-testen.

Omvendt er der ikke nogen der, da der bliver spurgt til det, angiver at have diskuteret nogle af de spørgsmål der indgår i gentestningen med studiekammerater eller lignende. Karakteristisk er for eksempel Tobias, der, da han præsenteres for spørgsmål 21, siger: *"Ja, det er faktisk et ret interessant problem, fordi jeg kan huske, at vi har siddet og lavet noget i denne her stil og jeg er stadigvæk ikke sikker på hvordan det er."* Dette opsummerer situationen ganske udmærket: rumskibsspørgsmålet har fæstnet sig i hukommelsen hos informanterne, og det, at spørgsmålet er bekendt kan have en indflydelse på, hvordan der svares. Omvendt er der ikke noget der tyder på, at informanterne aktivt har diskuteret eller søgt at løse spørgsmålene mellem FCI og gentestningen. Det er da også karakteristisk, at informanterne ikke altid kan genkalde, hvad de svarede i FCI. Eksempelvis siger Mette da hun får at vide, at alle spørgsmålene er nogen hun har set til FCI-testen, at *"så er det jo spændende at se om jeg svarer det samme nu som jeg gjorde den gang, det er jeg ikke sikker på"* og Søren angiver endda at

have svaret det samme i gentestningen som i FCI selvom hans besvarelse af både spørgsmål 13 og 30 faktisk er forskellig i de to tests.

Den samme gentestningsproblematik gør sig gældende i VAP-projektet, hvor der argumenteres for, at tæppebombningen af eleven med en stor mængde spørgsmål på ganske kort tid gør, at spørgsmålene ikke når at flytte sig fra korttids- til langtidshukommelsen hos eleven. En lignende tæppebombning må siges at finde sted i FCI-testen, hvor de studerende skal besvare 30 spørgsmål på 30 minutter. I (Dolin & Krogh, 2011, s. 45) konkluderes det på en statistisk vurdering af gentestningseffekter i VAP-resultaterne, at denne effekt er så lille at den kan ignoreres. Selvom der naturligvis er forskelle på VAP-projektets empiri og den empiri der er indsamlet i dette speciale, indikerer denne konklusion, at gentestningseffekterne næppe er altafgørende i forhold til resultaterne der observeres i den indsamlede empiri. Konklusionen bliver dermed, at der næppe er belæg for at tro, at de meget markante ændringer af besvarelserne kan tilskrives denne gentestningsproblematik. Det kunne dog være interessant at gennemføre en gentestning af en kontrolgruppe af studerende, der fik stillet de samme 6 spørgsmål i FCI's oprindelige evalueringsformat for at se, hvilke ændringer af besvarelserne det ville forårsage.

## 10.9 Konklusion

Vi har altså set, at der sker nogle meget grundlæggende ændringer af informanternes besvarelser når de gentestes i en sociokulturelt orienteret samtale. Det bliver derfor nødvendigt at forstå, hvad disse ændringer betyder for opfattelsen af FCI-resultater, hvilket vi skal vende tilbage til i kapitel 12. Herudover peger analysen på, at de studerendes forkerte besvarelser i FCI-testen ikke nødvendigvis indikerer tilstedeværelsen af stabile misconceptions. Dette får os til dyberegående at analysere et enkelt af spørgsmålene, nemlig spørgsmål 21, læringsteoretisk.

## 11 Læringsteoretisk analyse af den udførte empiri

Det ses altså, at gentestningen i et andet evalueringsparadigme i mange tilfælde ændrer de studerendes besvarelser af FCI-spørgsmålene. Dette kan som forklaret ovenfor forstås som et udtryk for, at det dialogbaserede, sociokulturelt orienterede evalueringsformat giver adgang til andre dele af informanternes viden end FCI's oprindelige evalueringsformat gør. Således er FCI analyseret ud fra et evalueringsteoretisk grundlag. Der kan ligeledes være behov for at forstå FCI ud fra et læringsteoretisk grundlag med udgangspunkt i den udførte empiri. Dette vil i det følgende blive gjort via analysen af et enkelt, særlig interessant spørgsmål, nemlig spørgsmål 21.

Spørgsmål 21 i FCI omhandler som bekendt et rumskib, der bevæger sig med jævn hastighed og påvirkes af en konstant kraft vinkelret på hastighedens retning. For at svare på spørgsmålet skal man afgøre, at rumskibet vil bevæge sig i en parabelformet bane. Denne situation er fysisk set helt ækvivalent til situationen i spørgsmål 12, hvor en kanonkugle skydes vandret af sted. Efter affyringen har kanonkuglen en konstant hastighed i affyringsretningen og påvirkes af en konstant tyngdekraft vinkelret på denne retning, hvilket vil føre til, at kanonkuglen bevæger sig i en parabelformet bane. Begge baner kan altså beskrives som skrå kast med en starthastighed der kun har en vandret komponent. Der er en betragtelig forskel på, hvor mange studerende der svarer korrekt på de to spørgsmål i FCI-testen. Data ses i et skema herunder:

Procentdel korrekte	Spørgsmål 12	Spørgsmål 21
2008	83%	62%
2009	88%	60%
2010	94%	66%
2011	90%	69%

Figur 13 Procentdel korrekte svar i spørgsmål 12 og 21

Der er altså mellem 21 og 29 procentpoints forskel på, hvor mange der svarer korrekt på de to spørgsmål, hvilket indikerer, at konteksten<sup>14</sup> i opgaven med kanonen på en eller anden måde er lettere eller mere genkendelig for de studerende end konteksten i rumskibsopgaven. Under udarbejdelsen af samtalskemaet blev det derfor besluttet, at studerende der ikke umiddelbart kunne tegne banen for rumskibet i spørgsmål 21 skulle stilles kanonspørgsmålet (spørgsmål 12) som et udgangspunkt for at diskutere deres forståelse af det skrå kast i en måske mere genkendelig kontekst. Også i interviewene var forskellen på forståelsen i de to situationer slående og da vi her har en indsigt i de studerendes ræsonnementer undervejs er det muligt at undersøge, hvorfor de

<sup>14</sup> I denne analyse anvendes begrebet kontekst helt snævert som en betegnelse for den konkrete repræsentation af et abstrakt fysisk koncept og må således ikke forveksles med for eksempel den sociale kontekst som diskuteres i kapitel 10.3. Konceptet i spørgsmål 12 og 21 er således det samme, en parabelbane forårsaget af en konstant kraft vinkelret på en konstant hastighed, mens konteksten er forskellig, nemlig henholdsvis en kanonkugle, der påvirkes af en konstant tyngdekraft og et rumskib der påvirkes af en konstant kraft fra sin motor.

to situationer opfattes så forskelligt af de studerende. Informanterne er i den følgende analyse inddelt i 4 forskellige grupper efter deres besvarelse af spørgsmålene.

### 11.0.1 Spørgsmål 21 besvaret korrekt og fyldestgørende

Under interviewene svarede fire studerende (Anna, Søren, Sarah og Mette) fuldstændig korrekt på spørgsmålet om rumskibets bane i den forstand at de både var i stand til at angive den korrekte bane og give en fyldestgørende fysisk forklaring på årsagen til, at netop denne bane følges. Eksempelvis svarer Anna således da hun bliver bedt om at angive rumskibets bane:

*Anna: For det første, hvis den er vinkelret på, så vil den stadig have sin komponent i den retning som den fløj i til at starte med og der vil den beholde sin jævne hastighed. Hvis den så har en kraft der påvirker, så vil det sige at den bliver accelereret og så ville accelerationen jo så, hvis den er konstant, jamen ville den jo umiddelbart.. så ville jeg sige at den foretog en eller anden form for parabelbane, halv parabelbane udad sådan her (demonstrerer med rumskibet).*

De vigtigste elementer i forklaringen, nemlig at vi har en konstant hastighed og en konstant acceleration vinkelret på hinanden og at dette giver anledning til en parabelbane er klart til stede.

Disse tre studerende blev således slet ikke stillet kanonspørgsmålet og er derfor kun relevante for denne analyse som eksempler på studerende der umiddelbart har en forståelse i rumskibets opgavens kontekst.

### 11.0.2 Spørgsmål 21 besvaret korrekt, men ikke fyldestgørende

En anden studerende (Peter) angiver umiddelbart den korrekte bane, og bliver derfor ikke stillet kanonspørgsmålet. Dog er forståelsen af, hvorfor denne bane er korrekt mere intuitivt baseret og formelt ikke fuldstændig korrekt. Peter siger således:

*Peter: Okay. Så vi har to kræfter. Eller den bevæger sig sådan her (bevæger rumskibet langs AB) og nu bevæger den sig så sådan her (bevæger rumskibet vinkelret på AB) og så er der jo min dejlige intuition igen der siger at den så må bevæge sig i noget der minder om en eksponentiel sådan den vej der (bevæger rumskibet i den korrekte bane).*

*Interviewer: Sådan buet kurve på en eller anden måde.*

*Peter: Buet denne her vej. Men det spiller vel også ind altså hvor stor den kraft der.. nu tegner du sådan en lille vektor der, hvor stor en kraft den bliver fyret af sted med. Hvis nu det var en million milliard, så ville den nok fyre rimelig meget ligeud.*

Vi ser altså en opdeling af bevægelsen i to komponenter, men ingen skelnen mellem den konstante hastighed og den konstante acceleration i de to forskellige retninger, muligvis fordi der er en grundlæggende forvirring omkring kraftbegrebet der blandes sammen med både hastigheds- og accelerationsbegrebet. Det er dog interessant, at Peter selv bringer det skrå kast i form af et eksempel med en bold på banen da han bliver bedt om at uddybe, hvorfor banen lige præcis har den form:

*Peter: Ja, det er fordi vi har to vektorer, som vi lægger sammen, og denne her bevæger sig med en vektor der peger denne her vej (langs AB) og så er der en der peger denne her vej (vinkelret på*

AB). Og så går den bare sådan lidt op. Altså, den vil gerne bevæge sig begge veje samtidig, kan man sige. Og lige præcis... Øh, jo hvis vi har de to kræfter.. Det er lidt ligesom når man sender en bold af sted, så laver den også den der parabel. Måske det er fordi.. Altså ting bare bevæger sig i parabler, det ved jeg ikke.. Det må jeg indrømme.

På et intuitivt, ikke-formaliseret niveau formår han altså at genkende de to situationer som manifestationer af den samme fysik og bygge bro fra en hverdagserfaring til den mere abstrakte rumskibssituation og vælger derfor den korrekte bane.

Nina befinder sig i en form for midterkategori. Hun er i starten noget tøvende omkring opgaven, men ræsonnerer sig efterhånden gennem diskussion med interviewerens frem til en korrekt forklaring:

*Nina: Hvis den går langsommere i starten, så ville den blive trukket mere mod højre før den nåede op, så buen ville gå den anden vej (som E'eren).*

Hun stilles efterfølgende også spørgsmål 12 og besvarer dette korrekt uden at hendes forklaring uddybes nærmere – hun angiver blot at, banen vil være parabelformet og der etableres ikke efterfølgende nogen forbindelse mellem de to spørgsmål. Det er interessant, at de to spørgsmål besvares på så forskelligt hos Nina. Spørgsmål 21 besvares således som kulminationen på en længere række af fysikfaglige overvejelser og ved hjælp af logiske ræsonnementer, mens besvarelsen af spørgsmål 12 er umiddelbar, intuitiv og let for Nina at genkalde. Hun virker altså til at være mere bekendt med den kontekst der optræder i spørgsmål 12.

Dermed er der altså 7 studerende tilbage som alle enten umiddelbart vælger den forkerte bane eller svarer så tilpas tøvende på spørgsmålet at de bliver stillet kanonspørgsmålet. Hvad der først og fremmest er karakteristisk er, at de alle 7 uden tøven tegner korrekte baner (med den usikkerhed der er forbundet med de hurtige frihåndstegninger). Kanoneksemplet er tydeligvis markant mere genkendeligt for dem på trods af, at der er tale om fysisk ækvivalente situationer og vi har tilsyneladende fat i noget centralt i forhold til, hvordan fysikken tillæres. Den fysiske forståelse af de samme koncepter er tilsyneladende større i nogle kontekster end i andre for disse 7 studerende. Der er dog, selvom kanonkuglens bane tegnes korrekt i alle tilfældene, forskel på, hvordan informanterne formår at forklare årsagen til, at netop denne bane følges.

### 11.0.3 Spørgsmål 12 besvaret korrekt og fyldestgørende

Fire af disse informanter (Tobias, Martin, Karoline og Lars) er i stand til at give en nogenlunde tilfredsstillende<sup>15</sup> formaliseret forklaring af årsagen til banens udformning, eksempelvis Tobias der forklarer kanonkuglens bevægelse således:

*Tobias: Det er sådan lidt det samme som vi havde i vores eksempel med bolden op og ned, at den bliver skudt fremad, og hvis vi så bare tænker formelmæssigt, så bevæger den sig med konstant*

---

<sup>15</sup> Det må bemærkes, at interviewerens ikke spørger særlig meget ind til forklaringen på boldens bevægelse. Det kan derfor have været svært for de studerende at vurdere, hvor grundig en forklaring der skal til for at svaret opfattes som fyldestgørende.

acceleration i x-aksens retning og i y-aksens retning der vil den begynde at falde nedad på grund af tyngdeaccelerationen og så rammer den jorden på et eller andet tidspunkt.

Interviewer: Okay. Nu skal jeg lige være sikker på, at du mener hvad du siger, du siger at den vil bevæge sig med konstant acceleration i x-retningen?

Tobias: Nej, hastighed i denne der retning.

Karoline forklarer bevægelsen således: "Altså, den har ikke nogen hastighed i y-retning, altså ikke op eller ned, den har kun en hastighed i x-retning, så den flyver ligeud for at begynde med. Men så virker tyngdekraften på kuglen, så bevægelsen vil bøjes nedad indtil kuglen lander."

Det kunne have været interessant at få en mere udfoldet redegørelse af kuglens hastighed og acceleration i de forskellige retninger, men da denne grundige analyse af forskellen i forståelsen af de to situationer ikke var planlagt, da empirien blev udført, blev der ikke spurgt ind til dette under interviewene, hvorfor ikke er muligt at vurdere de studerendes forståelse fuldt ud.

#### 11.0.4 Spørgsmål 12 besvaret korrekt, men ikke fyldestgørende

De resterende to studerende (Signe og Jakob) kommer med forklaringer hvis essens er korrekt (tyngdekraften påvirker kuglen således at den falder mod jorden), men som er af en noget vævende karakter. Eksempelvis beskriver Signe årsagen til kuglens bevægelse således:

Signe: Øhm... (Tegner en parabelbane) Noget i denne her stil vil jeg tro.

Interviewer: Ja, lige præcis. Det er helt rigtigt. Hvorfor vil den gøre det?

Signe: Fordi at... Den har en.. altså en fart, men så vil den blive påvirket af tyngdekraften og så vil det være ja, sådan der.

Der ses altså hos de adspurgte studerende en forskellig grad af forståelse af baggrunden for kanonkuglens bevægelse, men overordnet er ingen i tvivl om, at kuglen vil følge en parabelformet bane.

#### 11.1 Hvordan bygges bro mellem de to spørgsmål?

Vi ser altså, at 4 studerende i gentestningen er i stand at besvare et spørgsmål 12 både korrekt og tilfredsstillende uddybet selvom de ikke var i stand til at besvare spørgsmål 21 korrekt. Desuden er 2 studerende i stand til at besvare spørgsmål 12 korrekt, men ikke tilfredsstillende. I det følgende analyseres det, hvordan introduktionen af spørgsmål 12 påvirker informanternes besvarelse af spørgsmål 21 – ændres denne og hvordan anvender de i givet fald spørgsmål 12 til at forstå spørgsmål 21 med?

Lars formår faktisk selv at overføre forståelsen fra kanonopgaven til rumskibsopgaven. Han starter med at angive, at rumskibet efter motoren tændes vil bevæge sig i en ret linje, men da han bliver bedt om at genoverveje den oprindelige opgave set i lyset af kanoneksemplet gennemskuer han selv forbindelsen uden yderligere indblanding fra interviewerens:

Interviewer: (...) Får det dig til at tænke noget andet om rumskibets bevægelse herovre?  
(Stilhed 4 sekunder)

Interviewer: Eller vil det stadig bare bevæge sig sådan i denne her rette linje vi snakkede om før?  
(Stilhed 5 sekunder)

Lars: Der er en grund til at du spørger, går jeg ud fra. Øhm, hvad var det jeg sagde før, det var at den startede, så ville den bevæge sig sådan her (i en retlinet skrå kurve). Ja, det vil den blive ved

*med, for der ikke nogen sådan tyngdekraft der trækker den i nogen sådan bestemt retning. Nårh, prøv lige at vente lidt. Jo, okay, den vil bevæge sig sådan her (bevæger rumskibet i en buet bane).*

*Interviewer: Sådan lidt parabelformet.*

*Lars: Ja. For hvis man ser på de her raketmotorer som.. eller hvis vi ser bort fra raketmotorerne og forestiller os, at det var jorden det her og jorden var her (oven over raketten), så ville de blive trukket ned sådan, fordi det er jo sådan lidt det samme.*

Der er bemærkelsesværdigt, hvordan Lars laver denne forbindelse. Snarere end at betragte situationerne som to generelle eksempler på en uspecificeret konstant kraft vinkelret på en konstant hastighed oversætter han rumskibsopgaven til kanoneksemplet ved at indføre en tyngdekraft i stedet for motorkraften. Han er altså ikke i stand til at lave den generelle, abstrakte forbindelse, men laver i stedet forbindelsen via en konkret oversættelse af den opgave, han stilles, til en kontekst, der er tættere på det spørgsmål, han kan besvare. Dette viser, hvordan Lars' forståelse af den bagvedlæggende fysik er knyttet til forståelsen af eksempler, han har set tidligere. Der er altså ikke tale om en deduktiv slutning ud fra en grundlæggende teoretisk forståelse af Newtonsk mekanik, men derimod om anvendelsen af et standardeksempel på situationer der genkendes som ækvivalente.

Martin er et eksempel på en studerende der faktisk også i en eller anden forstand laver en oversættelse fra det ene spørgsmål til det andet, men uden det helt er tydeligt, hvad der får det til at ske. Han ræsonnerer således under interviewet:

*Interviewer: Så med de her 5 valgmuligheder, så sagde du at det var denne her (peger på C'eren)?*

*Martin: Ja.*

*Interviewer: Ja.*

*Martin: Og så skulle det så være.. Den bliver påvirket med en konstant kraft og vi slipper den i den retning. Så skulle den.. Ja. Ja, det kunne vel så godt være at jeg lige tog fejl der, tænker jeg, at den så bevæger sig.. ja, det må jo så være E, ikke. Den bliver påvirket af en konstant kraft, så den drejer (bevæger rumskibet i en parabelbane).*

Tilsyneladende laver han en forbindelse mellem de to spørgsmål der er mere intuitiv end den er formaliseret, om end han har fat i, at den konstante kraft er central, og udgør på den måde en slags midterkategori mellem Lars, der laver en formaliseret oversættelse og Signe og Jakob, der slet ikke laver oversættelsen.

Karoline er i udpræget grad ikke i stand til at aflæse den kontekst, som spørgsmål 21 foregår i, men tolker det, at opgaven foregår i det ydre rum som afgørende for, hvilket svar der er det korrekte. Hendes ræsonnement kan aflæses af følgende interviewuddrag:

*Karoline: (...) så tænkte jeg, at det er jo ude i rummet og måske kan man bare flyve sådan nogle retvinklede kurver i rummet.*

*Interviewer: Okay, fordi det godt kan være at tingene opfører sig anderledes i rummet end vi er vant til hernede på jorden?*

*Karoline: Ja, men hvis det.. altså hvis de opfører sig på samme måde så burde den også flyve sådan (bevæger rumskibet i en parabelbane).. ja, i sådan en kurve.*

Hun er efterfølgende i stand til at give en nogenlunde tilfredsstillende, om end ikke uddybende fysisk forklaring af parabelbanen. I den forstand har Karoline ikke som sådan misconceptions relateret til parabelbaner. Til gengæld er hun ikke i stand til at dekontekstualisere rumskibetsopgaven og dermed gennemskue, hvilke fysiske koncepter hun skal anvende for at løse opgaven tilfredsstillende, hvorfor hun alligevel svarer forkert. Noget tilsvarende ses i relation til Tobias. Som det er gennemgået i kapitel 10.4 har han nogle tekniske overvejelser i forhold til rumskibets acceleration, der gør, at han ikke vælger den korrekte bane. Efter en længere samtale og en vist mål af stilladsering viser det sig, at han ikke er i tvivl om, hvilken bane der er den korrekte hvis rumskibets acceleration i kraftens retning er konstant, men at han til gengæld har hverdagsforståelser der gør, at han ikke er overbevist om, at accelerationen er konstant. Igen er der måske ikke så meget tale om misconceptions omkring parabelbaner hos Tobias som der er tale om en manglende evne til at gennemskue konteksten og formå at anvende formelle fysiske ræsonnementer på en situation der forstås gennem en hverdagsoptik.

Hos Tobias og Karoline er det altså ikke så interessant at tale om, hvorvidt de formår at lave en oversættelse fra spørgsmål 12 til spørgsmål 21, al den stund at det der forhindrer dem i at vælge den korrekte bane ikke er deres forståelse af parabelbaner, men derimod en mangle evne til at gennemskue, at det er denne viden skal bringes i spil i disse situationer. Konteksten, for Karoline at opgaven foregår i rummet og for Tobias at der er tale om et fysisk rumskib der må formodes at have en tophastighed, har en betydning for, hvordan de er i stand til at gennemskue, hvilke fysik, der er relevant at anvende.

Signe og Jakob er ikke i stand til at komme tættere på en forklaring af spørgsmål 21 selvom de er i stand til at angive den korrekte bane for kanonkuglen. Jakob kan genkende den korrekte bane som magen til den bane han har tegnet for kanonkuglen, men kan ikke give en formaliseret forklaring af rumrakettens bane set i lyset af kanonkuglens bevægelse. Det er karakteristisk, at det netop er disse to studerende der har sværest ved at overføre deres kendskab til kanonkuglens bane til den nye situation. Deres kendskab til kanonkuglens bevægelse er kendetegnet ved, at de kan gengive karakteristika ved bevægelsen, men at de ikke er i stand til at ræsonnere fysikfagligt omkring den.

Opsummerende kan det altså siges, at det er vanskeligt for de studerende at opnå en formaliseret forståelse af de to situationer som ens. Det aktivt at blive bedt om at lede efter forbindelseslinjer mellem spørgsmålene kan dog forårsage at denne genkendelsesproces påbegyndes og eventuelt gennemføres som det ses for Lars og Martin. Derudover ses det, at opgavernes kontekst kan være afgørende for, om det er muligt at genkende to situationer som ens. Karolines manglende evne til at dekontekstualisere spørgsmålet hindrer hende i at genkende spørgsmål 21 som udtryk for samme fysik som spørgsmål 12, men hendes overvejelser giver et indtryk af, at hun ville være i stand til at gennemføre en oversættelse for et spørgsmål med samme konceptuelle indhold men en anden kontekstuel repræsentation.

### 11.2 Hvordan fortolkes ovenstående læringsteoretisk?

I det følgende søges ovenstående, forskellen på kunnen i relation til to forskellige kontekster, forstået ud fra FCI's læringsteoretiske grundlag. Efterfølgende gives der andre læringsteoretiske bud på en forståelse af forskellen.



Først og fremmest er det værd at overveje, i hvor høj grad de to spørgsmål kan siges at være udtryk for samme bagvedlæggende fysik.

I begge tilfælde er der tale om et objekt, der bevæger sig med en konstant vandret hastighed og en konstant lodret acceleration, enten tyngdeaccelerationen eller accelerationen forårsaget af den konstante kraftpåvirkning fra motoren, hvorfor den korrekte bane kan forklares ved anvendelse af præcis samme argumentation i begge tilfælde. Alligevel må man nok tale om en forholdsvis stor kontekstuel afstand mellem de to spørgsmål. Dette afspejles af, at de to spørgsmål faktisk placeres i hver sin konceptuelle dimension i (Hestenes et al., 1992). Således indplaceres spørgsmål 21 under kinematik og herunder i kategorien "Constant acceleration entails parabolic orbit", hvorimod spørgsmål 12 (om kanonkuglens bane) placeres under "parabolic trajectory" under gravitation, der igen er placeret under "kinds of force". Ligeledes afspejles det ved, at svarmuligheder til de to spørgsmål er forskellige. Således er banerne 1 og 2 i spørgsmål 21 ikke valgmuligheder i spørgsmål 12.

Når det alligevel giver mening at karakterisere de to spørgsmål som udtryk for den samme fysik også i FCI's forståelse, er det fordi, at de to spørgsmål har til hensigt at teste nogle af de samme misconceptions, nemlig "Force compromise determines motion" (svarmulighed 1 i spørgsmål 12 og 3 i spørgsmål 21) samt misconceptions omkring impetus. Således er det i FCI's forståelse to forskellige tests af samme misconception.

Efter altså at have fastslået, at vi grundlæggende må anse de to situationer som forskellige kontekstuelle repræsentationer af samme bagvedlæggende fysik, der intenderer at teste den samme misconception er det relevant at se, hvordan den observerede empiri kan forstås ud fra FCI's læringsteoretiske standpunkt. Helt overordnet må det konkluderes, at observationerne stemmer dårligt overens med en forståelse af, at er korrekt Newtonsk verdensbillede tilegnes i alle kontekstuelle repræsentationer inden for samme konceptuelle områder på en gang. Grundlæggende er det ikke meningsfuldt at tale om, at informanterne har en force compromise determines motion-misconception eller at de omvendt har konstrueret en Newtonsk forståelse af paraboliske baner. Snarere kan man tale om, at de studerende har en misconceptions der er kontekstafhængige – nogle kontekster får dem til at anvende en Newtonsk tankegang og andre gør ikke. Disse observationer er vanskeligt forenelige med FCI's læringsteoretiske, conceptual change-inspirerede grundlag som det er gennemgået i kapitel 5.6.1. Her ses der en forståelse af, at læring foregår ved at misconceptions, der ofte har rod i hverdagsforståelser, fortrænges af korrekte, Newtonske koncepter. Dette harmonerer ikke med observationerne af, at de studerende tænker Newtonsk i nogle kontekster og ikke i andre. Desuden er det interessant at overveje, hvordan forståelsen af, at de studerendes misconceptions bunder i hverdagsforståelser relaterer sig til det observerede. Der er næppe nogle af de to situationer, der kan siges at have en direkte forbindelse til hverdagen for de studerende – trods alt er hverken kanoner eller rumskibe noget man møder særlig tit. Snarere kan man måske tale om, at kanonspørgsmålet har en lille kontekstuel afstand til andre situationer, der kendes fra hverdagen (kast med bolde for eksempel), hvilket gør, at det opleves som mere genkendeligt ud fra en hverdagsforståelse. Dette nævnes da også for eksempel af Lars, der siger, at "*det her (kanonen) er jo sådan intuitivt fordi man kender fra hverdagen at når man kaster noget at det bevæger sig sådan der, ikke*". I forhold til forståelsen af, at misconceptions bunder i hverdagsforståelser og dagligdags erfaringer er det derfor bemærkelsesværdigt, at den situation der opleves som hverdagsagtig faktisk er den, der får de studerende til at tænke Newtonsk, mens den mindre genkendelige rumskibsopgave giver anledning til, at ikke-Newtonske misconceptions kommer i spil.

Noget kunne således tyde på, at der er behov for en mere nuanceret læringsteoretisk forståelse af, hvordan de studerendes Newtonske forståelse måles i FCI, således at billedet af, at studerendes Newtonske tænkning er kontekstafhængig inkorporeres.

En dybdegående, nuanceret gennemgang af, hvordan disse empiriske observationer kan forstås ud fra andre læringsteoretiske udgangspunkter er uden for rammerne af dette speciale. Det følgende skal altså ses som et udgangspunkt for en nærmere diskussion af, hvordan FCI-resultaterne kan forstås læringsteoretisk, snarere end som en fuldgyltig diskussion i sin egen ret.

Man kan overveje, hvad det kan skyldes, at den ene kontekst, kanonen på en bakketop, er markant mere genkendelig for de studerende end konteksten i rumskibsopgaven. For det første kan det som tidligere nævnt skyldes, at kanonopgaven lettere kan relateres til situationer, som de studerende har konkrete, muligvis også kinæstetiske, erfaringer med. Herudover er det sandsynligt, at kanonspørgsmålet er genkendeligt fordi det udgør hvad man kunne kalde et arketyrisk eksempel eller problem i den klassiske mekanik.

Dette kan forstås ud fra Thomas Kuhns ide om, at videnskabelige paradigmer karakteriseres ved at have en samling af særlige problemstillinger og tilhørende løsninger, såkaldte eksemplarer, som er med til at udgøre feltet. Dette kunne i fysikken for eksempel være skråplanet, den harmoniske oscillator eller det skrå kast. At lære fysik handler om at blive i stand til at anvende disse eksemplarer til at forstå andre situationer som andre fremstillinger af det samme fysiske problem. Således hedder det i (Kuhn, 2012, postskript):

*“The resultant ability to see a variety of situations as like each other, as subjects for  $f = ma$  or some other symbolic generalization, is, I think, the main thing a student acquires by doing exemplary problems (...). After he has completed a certain number, which may vary widely from one individual to the next, he views the situations that confront him as a scientist in the same gestalt as other members of his specialists' group.”*

Herudfra kan man betragte observationerne som et udtryk for, at de studerende har kendskab til og forståelse for et arketyrisk problem eller eksemplar inden for den Newtonske mekanik, det vandrette kast, men at kun enkelte af dem formår at genkende dette eksemplar når det optræder med en for stor kontekstuel afstand til den form, eksemplaret oprindeligt er præsenteret for dem i. At lige præcis konteksten i spørgsmål 12 er genkendelig som en kontekst eksemplaret tidligere har været præsenteret for de studerende kan meget vel skyldes, at ballistik tidligere har været et helt centralt område, mekanikken har været anvendt på og udviklet i.

I relation hertil må det dog bemærkes, at brugen af Kuhns videnskabsteoretiske og -historiske værk som læringsteori ikke bør strækkes alt for langt. Ideen om eksemplarer inden for et videnskabeligt felt giver en forståelse af, hvorfor det ene spørgsmål er så markant lettere for de studerende, men tilbyder ikke meget forståelse af, hvordan læringen faktisk finder sted. Således betragtes læring som en simpel stimuli-respons-proces – med et citat lærer man *“to see the same things when confronted with the same stimuli (...) by being shown examples of situations that their predecessors in the group have already learned to see as like each other and as different from other sorts of situations”* (Kuhn, 2012, postskript). Populært sagt kan det ses som den samme proces der finder sted, når forældre gentagne gange udpeger en hund og barnet til sidst selv bliver i stand at genkende en hund når det ser en.

Denne forståelse harmonerer dog ikke særlig godt med den udbredte forståelse af, at det generelt kan være vanskeligt at lave denne såkaldte transfer, altså at anvende den viden man har om en bestemt kontekst på andre kontekster (Dolin, 2006, s. 157). En læringsteoretisk

forståelse af, hvordan de studerende forstår de to spørgsmål bør således tage hensyn til, at der også er brug for at forstå, hvordan de studerende formår at genkende eller netop ikke at genkende de to spørgsmål som de samme. At forståelsen af disse transferprocesser spiller en væsentlig rolle for forståelsen af, hvordan resultaterne af FCI-testen skal forstås læringsteoretisk påpeges også i (Huffman & Heller, 1995a), som det blev gennemgået i kapitel 5.5. I (Royer, Mestre & Dufresne, 2005) påpeges det, at muligheden for at gennemføre denne transferproces er influeret af en del forskellige forhold. Nærmere diskussioner af, hvilke forhold der i litteraturen er belæg for har en indflydelse på, hvordan transferen foregår skal ikke ske her. Vi skal blot bemærke, at denne forståelse af, at transfer af viden ikke altid vil ske, er beskrivende for den observerede empiri. Her kan der ske en transferproces, således at viden om spørgsmål 12 anvendes på spørgsmål 21, ved at informanten bliver cuet til at lede efter forbindelser mellem de to situationer, hvilket ses hos Lars og til en vis grad hos Martin. Evnen til at afdække og gennemskue kontekster kan dog hindre denne proces som det ses hos Karoline og Tobias, ligesom eksemplerne med Signe og Jakob, der ikke foretager en transfer af viden fra den ene opgave til den anden, peger i retning af, at graden af forståelsen af spørgsmål 12 hænger sammen med, hvorvidt det er muligt at gennemføre transferprocessen. (Huffman & Heller, 1995a) forklarer også de observationer der er gjort ud fra en forståelse af de studerendes betydningsdannelse i relation til de forskellige spørgsmål som aktiveringen af forskellige p-prims hos den studerende. Et sådant læringsteoretisk udgangspunkt kan også bidrage til en forståelse af, hvad der er observeret ovenfor. Helt kort kan det siges, at den studerende besidder p-prims, der aktiveres til at konstruere en Newtonsk korrekt forståelse i spørgsmål 12. I spørgsmål 21 aktiveres i stedet andre p-prims, hvilket gør, at den studerende ikke forstår dette spørgsmål korrekt Newtonsk. Der er således ikke tale om, at den studerende har forkerte eller ukorrekte p-prims, men i stedet, at det er uhensigtsmæssige p-prims der aktiveres i spørgsmål 21. En nærmere diskussion af, hvilke p-prims der aktiveres og hvordan dette sker er dog uden for rammerne af dette speciale. Blot er det væsentligt at bemærke, at vi også her ser et bud på en mentalistisk orienteret læringsteori der kan forklare det observerede.

### 11.3 Konklusion

Vi har altså set, at der er en markant forskel på studerendes viden i to forskellige kontekstuelle repræsentationer af samme fysiske koncept.

Disse observationer stemmer dårligt overens med den læringsteoretiske forståelse der ligger til grund for FCI-testen og som tager udgangspunkt i en forståelse af, at forkerte svar i FCI-testen er et udtryk for bagvedlæggende misconceptions. Dette betyder, at der er behov for at overveje andre læringsteoretiske perspektiver på FCI-testen og vi har set, at der findes andre mentalistisk orienterede syn på læring, der kan forklare den observerede forskel på besvarelsenerne og som altså giver andre perspektiver på de læringsprocesser der ligger til grund for den viden der evalueres i FCI-testen.

## 12 Diskussion

Dette speciale har udviklet sig fra først og fremmest at beskæftige sig med kønsforskellen i FCI-testen til i stedet at undersøge denne test evaluerings- og læringsteoretisk. Følgende afsnit skal tjene som en diskussion af, hvilken betydning disse resultater har for forståelsen af FCI og hvordan denne kan anvendes.

### 12.1 Evalueringsteoretisk diskussion

Det påpeges i VAP-projektet, at *"Ud fra en postpositivistisk tilgang ville man nok sige, at elevernes bedre præstationer skyldes at de er blevet hjulpet, men i en sociokulturel sammenhæng vil man sige at den kommunikative situation er ændret og dermed elevens kunnen"* (Dolin & Krogh, 2012, s. 87). Fortolkningen af resultaterne afhænger altså af det evalueringsparadigmatiske ståsted man betragter testningen fra. At man betragter resultataendringen som et validt udtryk for elevernes kompetencer handler om hvorvidt man vil erklære sig enig i det grundlæggende evalueringsparadigme og dets baggrund i sociokulturel læringsteori og dertilhørende epistemologi.

Der er dog ikke nogen tvivl om, at informanternes besvarelser ændres drastisk når man får indsigt i dem via den sociokulturelt orienterede samtale og ikke gennem FCI-testen med dens multiple choice-format. Således må det konstateres, at billedet af den enkelte studerende og hans eller hendes konceptuelle forståelse i høj grad er afhængigt af, hvordan man vælger at evaluere den.

FCI's multiple choice-format har nogle klare fordele. Først og fremmest gør det det muligt at teste mange studerende – gerne hele årgange – med et overskueligt tidsforbrug til både selve testningen og resultatgivningen. At teste hele årgange ved hjælp af sociokulturelt orienterede samtaler er så tidskrævende, at det må betragtes som urealistisk. Herudover er resultaterne af FCI let sammenlignelige på tværs af uddannelsesinstitutioner og årgange idet testen har en høj grad af reliabilitet. Opgaverne ændres ikke og der er ingen diskussion af, hvordan de skal pointgives. I en postpositivistisk forståelse kunne man således tale om, at FCI er et objektive måleinstrument der gør det muligt at vurdere effekten af undervisning eller at sammenligne studerendes forståelse på tværs af årgange. I en sociokulturel forståelse er ideen om eksistensen af en sådan objektiv evalueringsform dog meningsløs.

Jeg vil ikke anfægte, at denne reliabilitet og mulighed for sammenligning er væsentlig hvis man ønsker at bruge FCI til for eksempel at vurdere, om der er forskel på forskellige årgange og deres forståelse af klassisk mekanik. FCI-scoren er et mål for denne forståelse. Men samtidig viser resultaterne af dette speciale også, at dette mål er knyttet sammen med testens udformning og hjemsted i det givne postpositivistiske evalueringsparadigme. FCI-scoren giver dermed ikke nødvendigvis et entydigt eller udtømmende billede af hvad den enkelte studerende kan – billedet både nuanceres og ændres i den sociokulturelt orienterede samtales udvidede opmærksomhedsvindue.

Jeg vil mene, at opmærksomhed omkring dette er væsentligt – både generelt, men måske særligt i relation til brugen af FCI som diagnostisk værktøj der kan bruges til at udtale sig om den enkelte studerende på basis af FCI-scoren. Når FCI således bruges absolut og ikke som en sammenligning mellem større grupper af studerende eller sammenligning af den enkelte studerendes præ- og postscore i relation til et givet undervisningsforløb, er det væsentligt at

være bevidst om, at dette mål ikke er mere absolut eller objektivt end at det er farvet af FCI's evalueringsformat og evalueringsparadigmatiske ståsted.

At det er nødvendigt at være bevidst om FCI's evalueringsparadigmatiske baggrund er endvidere relevant idet testen søger at indfange resultatet af nogle læringsprocesser der har fundet sted på de gymnasiale uddannelser og altså dermed i en anden kontekst end testen foretages i. Således hedder det i (Dolin & Krogh, 2012, s. 87) at *"Størst udsigelseskraft må resultaterne anses at have, når den valgte evaluering er i overensstemmelse med det paradigme og de værdier, som lå til grund for elevernes læreprocesser."* En lærings- og evalueringsteoretisk analyse af de gymnasiale læreplaner er uden for rammerne af dette projekt, men den potentielle uoverensstemmelse kan have en betydning for, hvordan FCI-resultaterne skal forstås – ligesom det peger på, at en direkte sammenligning af FCI-scoring mellem elever, hvis læreprocesser i relation til klassisk mekanik er foregået med udgangspunkt i forskellige læringssyn ikke er simpel.

I kapitel 5.4 diskuteres FCI's validitet og reliabilitet som den er etableret i (Hestenes et al., 1992) og (Halloun & Hestenes, 1985a). Et aspekt af denne validitet er etableret på baggrund af interviews, hvis metodiske baggrund ikke er nærmere diskuteret, i hvilke de interviewede studerendes besvarelser af FCI-spørgsmål altovervejende er i overensstemmelse med deres besvarelser i den skriftlige FCI- eller MD-test. Konklusionerne i dette speciale afviser ikke endegyldigt at FCI-testen er valid i dette aspekt af validitetsbegrebet, men resultaterne af gentestningen udfordrer denne validitet i en dansk kontekst. Det mål for studerendes Newtonske tænkning som FCI-testen udgør er ikke mere entydigt end det kan ændres ved en ændring af det evalueringsparadigmatiske ståsted som evalueringen gennemføres under.

## 12.2 Læringsteoretisk diskussion

Som nævnt ovenfor måler FCI-testen studerendes forståelse af klassisk mekanik. De aspekter af FCI's validitet som omhandler, hvorvidt spørgsmålene dækker det Newtonske kraftbegreb bredt og tilstrækkelig udtømmende, ændres ikke af de konklusioner der drages i dette speciale. Til gengæld er det relevant at diskutere hvad dette begreb, forståelsen af klassisk mekanik, dækker over. Som det ses i 5.6.1 er det læringsteoretiske udgangspunkt for FCI-testen mentalistisk præget og baserer sig på, at studerende har hverdagserfaringer, der giver dem misconceptions som den vellykkede undervisning adresserer og udrydder.

Først og fremmest stiller den evalueringsteoretiske analyse spørgsmålstegn ved, om denne forståelse kan betragtes udelukkende ud fra et mentalistisk perspektiv. Et sociokulturelt perspektiv der påpeger betydningen af sproget som artefakt i elevernes meningsdannelse giver perspektiver på forståelsen af forskellen på besvarelserne i de to evalueringsparadigmer som et rent mentalistisk blik ikke rummer. Således har vi brug for begge perspektiver i forståelsen af gentestningens resultater.

Herudover har analysen af besvarelserne af spørgsmål 21 vist nogle sider af elevernes forståelse, der vanskeligt fanges af læringssynet som det er diskuteret i kapitel 5.6.1 – det er vanskeligt at forstå informanternes vanskeligheder omkring besvarelserne af dette spørgsmål udelukkende som et spørgsmål om, hvorvidt de har bestemte misconceptions. Snarere er deres misconceptions kontekstbundne – iscenesættelsen af det fysiske koncept i bestemte kontekster hvor kræfterne, objekterne og omgivelserne varierer præger, hvordan de studerende tænker omkring fysikken i opgaven. Dette betyder, at man måske skal se FCI ikke så meget som en måling af studerendes misconceptions men snarere som en måling af hvor bredt de studerende formår at anvende de korrekte Newtonske koncepter. At en studerende

svarer forkert på et FCI-spørgsmål er således ikke et udtryk for, at misconceptionen som de kan aflæses i Table II i Appendix B er gennemgående og karakteristisk for den studerendes forståelse, men blot at den eksisterer i netop det spørgsmåls kontekst. Ligeledes er tilstedeværelsen af Newtonske koncepter i bestemte sammenhænge ingen garanti for, at de kan anvendes i andre kontekstuelle sammenhænge. Dette peger også på, at en for skarp inddeling af studerendes forståelse som enten Newtonsk eller ikke-Newtonsk er unuanceret. Dette antyder, at den i visse fysikdidaktiske sammenhænge udbredte læringsteoretiske opfattelse af hverdagsbaserede misconceptions, der erstattes af videnskabeligt etablerede koncepter, ikke altid er et tilfredsstillende billede på læringsprocesser i forbindelse med Newtonsk mekanik. Disse konklusioner ligger i forlængelse af de konklusioner der drages i (Huffman & Heller, 1995a), selvom den metodiske baggrund i denne artikel er markant anderledes end den metodik der er anvendt i dette speciale. I kapitel 11.2 gives der bud på, hvordan disse læringsprocesser kan forstås ud fra andre læringsteoretiske udgangspunkter. Grundlæggende må det dog nok konstateres, at en uddybet og velundersøgt forståelse af, hvordan disse læringsprocesser som FCI måler resultatet af er uden for rammerne af et speciale af denne størrelse. Specialet peger dermed i retning af, at nærmere undersøgelser af, hvordan FCI skal forstås læringsteoretisk er ønskeligt.

## 13 Perspektivering

I det følgende gives der et bud på, hvordan den udførte empiri og de konklusioner der drages i specialet kan bruges til at tilrettelægge undersøgelser, der udforsker de opstillede forskningsspørgsmål nærmere.

### 13.1 Kønsforskellen i FCI

Et væsentligt mål i udarbejdelsen af denne opgave har været at undersøge, hvorvidt kønsforskellen i FCI-testen, der er observeret i udlandet, genfindes i en dansk kontekst og i givet fald søge at forstå, hvad denne kønsbias skyldes.

Det er næppe realistisk at tro, at der kun er en enkelt forklaringsmodel der gør sig gældende i forhold til denne kønsforskel. Snarere er forståelsen af den temmelig komplekse problemstilling bundet op på forståelsen af et sammensat billede af flere forskellige faktorer der spiller ind på den observerede forskel.

Dette speciale har primært beskæftiget sig med forklaringsmodeller, der handler om at søge at afdække en eventuel indbygget kønsbias i selve testen, hvad enten det handler om en bias i evalueringsformatet eller om en bias i form af ikke-inkluderende opgavekontekster. Selv om det ikke definitivt er afvist, at kønsforskellen ville tage sig anderledes ud, hvis evalueringsformatet ændredes helt grundlæggende, er der dog ikke ud fra de undersøgelser, der er foretaget i specialet, belæg for at tro, at hele kønsforskellen bunder i selve udformningen af testen.

Det er derfor værd at overveje, hvilke andre forklaringsmodeller der kunne undersøges for at blive klogere på den observerede kønsforskel.

Først og fremmest kunne det være relevant at undersøge, hvorvidt kønsforskellen observeres andre steder, som jeg ikke har haft adgang til data omkring. Er der for eksempel flere kvinder en mænd der har haft fysik på A-niveau når de optages? Har de optagne mænd i gennemsnit højere karakter i fysik på gymnasieniveau end de optagne kvinder? Vi har set, at kvinderne har højere gennemsnitlige karakterer, men hvis mændene til gengæld har højere karakter i fysik kan en del af kønsforskellen måske forklares ved dette. Ligeledes kunne det være relevant at undersøge, om den samme kønsforskel genfindes i posttestresultaterne eller om universitetsundervisningen ændrer på kønsforskellen.

I forhold til forklaringsmodeller der baserer sig på en forståelse af, at FCI-resultaterne faktisk afspejler en eksisterende forskel i de to køns forståelse af den Newtonske mekanik er der to, jeg vil pege på som interessante at undersøge.

Man kunne forestille sig, at en høj grad af kinæstetisk eller kropslig forståelse af den fysik der optræder i FCI-testen kunne resultere i en højere score. Således kunne en hypotese være, at drenge præsterer bedre i FCI som følge af en højere grad af intuitiv, kinæstetisk forståelse af mekanik som er erhvervet gennem for eksempel fysisk leg og sport og som altså er tillært uden for de formelle læringsmiljøer. Eksempelvis taler Peter i interviewet om, hvordan hans erfaringer fra tennisspil fortæller ham, at der må virke en luftmodstand på bolden i spørgsmål 30.

Væsentligst vil jeg dog mene at det er, at der i (Hestenes et al., 1992) peges på, at der er en sammenhæng mellem attitude, i forstanden forestillinger om faget fysik og hvilke

læringsstrategier man mener er hensigtsmæssige, og FCI-score. Denne attitude undersøges i tests som for eksempel den modererede udgave af The Maryland Physics Expectation (MPEX) survey der er anvendt i (Nielsen, 2004). I denne undersøgelse hvor man skal angive sin enighed i udsagn som *"Den bedste måde at lære fysik for mig er ved at løse mange problemer frem for at analysere få problemer i dybe detaljer"* eller *"I dette kursus regner jeg ikke med at forstå ligningerne på en intuitiv måde. De skal bare tages for givet."* I (Nielsen, 2004) undersøges sammenhængen mellem besvarelser i FCI og den modererede MPEX-undersøgelse for fysikstuderende ved Københavns Universitet og der ses en tendens til at kvindernes besvarelser i MPEX-testen udtrykker en mere u hensigtsmæssig attitude til faget end mændenes. Denne sammenhæng kunne være værd at udforske nærmere for at blive klogere på kønsforskellen i FCI.

Det kunne være interessant at se, om en bevidstgørelse om hensigtsmæssig faglig attitude hos de studerende kunne mindske kønsforskellen i FCI, ligesom det kunne være værd at analysere, hvordan de lavere MPEX-scoringer hos kvinderne kunne tænkes at hænge sammen med, hvordan de simultant konstruerer en kønsidentitet og en identitet som fysikstuderende i mødet med det kønsskæve studium.

Selvom specialets fokus undervejs i udarbejdelsen har forskudt sig fra primært at beskæftige sig med FCI ud fra et kønsperspektiv til snarere at betragte FCI i et evaluerings- og læringsteoretisk lys, vil jeg mene, at den særdeles markante kønsforskel i FCI er så overraskende set i lyset af, at kønsforskellen hverken genfindes i de adgangsgivende karakterer eller i de første universitetskarakterer, at videre bestræbelser på at forstå denne kønsforskel er både nødvendige og ønskelige.

### 13.2 Videre anvendelser af den indsamlede empiri

Et stort og væsentligt resultat i udarbejdelsen af dette speciale har været indsamlingen af den betragtelige mængde empiri. Denne empiri kan analyseres grundigere end rammerne for specialet tillader, hvorfor der i det følgende vil blive givet bud på, hvordan man fremadrettet kan anvende empirien.

Først og fremmest giver empirien et indblik i et bredt udvalg af studerendes ræsonnementer og fysikfaglige argumentation i relation til et område af den klassiske mekanik. Interviewene giver derfor mulighed for et mere nuanceret indblik i nystartede fysikstuderendes måde at tænke klassisk mekanik på end FCI-resultaterne umiddelbart giver adgang til. Dette er særlig interessant, fordi sådanne undersøgelser af studerendes forståelser af Newtonsk mekanik tidligere typisk er gennemført i en amerikansk kontekst. Således giver empirien mulighed for at undersøge, om der er bestemte faglige områder eller ræsonnementer der er særligt vanskelige for de studerende. Således er der for eksempel noget der tyder på, at de studerendes forståelse af, hvornår og hvorfor man tillader sig at se bort fra luftmodstanden ikke altid er tilfredsstillende. Denne forståelse kunne udforskes nærmere ud fra den indsamlede empiri.

Empirien kan også bruges til at få et dybere indblik i de studerendes måde at anvende sproget som artefakt på, end pladsen i dette speciale umiddelbart tillader.

Således kan det blandt andet undersøges, i hvor høj grad de studerende anvender hvad man kunne kalde et matematiseret sprog til at besvare opgaverne med og om der eksisterer en sammenhæng mellem denne evne til at anvende et matematisk sprog og for eksempel tale om vektorer, som for eksempel hastighederne  $v_y$  og  $v_x$ , og den konceptuelle mekaniske forståelse. En sådan undersøgelse er særlig interessant idet FCI jo netop er fritaget for krav om



matematisk forståelse. Alligevel ses der en forskel i anvendelsen af matematiske termer hos informanterne og det kunne være interessant at overveje, om denne anvendelse af sproget som et artefakt til at udtrykke sin forståelse gennem er korreleret med den konceptuelle forståelse.

Ligeledes er det ud fra empirien muligt at undersøge, hvordan den usikkerhed, der i FCI-testen kom til udtryk som udeladelser af svar kom til udtryk i det nye evalueringsformat. De studerende udtrykte undervejs tvivl om hvad de lavede, både begrundet hvor der tilsyneladende var en overensstemmelse mellem den udtrykte tvivl og den faglige forståelse, men indimellem også mere ubegrundet som når for eksempel Nina efter at have besvaret spørgsmål 21-24 korrekt alligevel modererer det ved følgende udsagn:

*"Se, det er meget bedre end ketsjere, det her. Jeg ved ikke om det passer, men det er.."* Også i forhold til overvejelserne omkring kønsforskellen i FCI kunne det være interessant at overveje, om der ses en kønsforskel i, hvordan der udtrykkes tvivl undervejs i interviewene.

Interviewene indeholder desuden en afsluttende del, som kun i begrænset grad er blevet analyseret. Det primære formål med denne afsluttende samtale var at få mulighed for at vurdere, i hvor høj grad informanterne drog fordel af, at den sociokulturelle evaluering var en gentestning af spørgsmål de havde besvaret tidligere. Denne del af interviewet var væsentligt mindre styret og udviklede sig efterhånden som samtalen udfoldede sig. Dette betød, at mere generelle træk ved de studerendes opfattelse af FCI-testen kom frem, ligesom det gav et indblik i hvordan de studerende opfatter klassisk mekanik som fagområde (for eksempel hvorvidt de forholder sig til, om deres hverdagserfaringer kan bruges til at forstå fysikken i formaliserede læringssituationer). En nærmere analyse af for eksempel hvad de studerendes opfattelse af, hvad formålet med at bliver stillet FCI-testen er, kunne give interessante perspektiver på for eksempel studiestarten og hvordan de studerende i løbet af denne påbegynder en konstruktion af en identitet som fysikstuderende.

## 14 Konklusion

Formålet med dette speciale har været følgende:

- At undersøge, hvorvidt den i udlandet observerede kønsforskel genfindes på NBI og at undersøge hvilken rolle de studerendes udeladte svar har i forhold til FCI-resultaterne og at undersøge, om der ses en kønsforskel her der kan være med til at forklare den observerede kønsforskel i FCI.
- At undersøge, hvad der sker med resultaterne af FCI-testen hvis evalueringsformen ændres fra en traditionel multiple choice-test til en sociokulturelt orienteret evaluering og at forstå, hvad denne resultatændring skyldes.

Det første forskningsspørgsmål er besvaret ved en kvantitativ dataanalyse af resultater fra FCI-testen fra årene 2008 til 2011.

Det viser sig, at der er en signifikant forskel på de gennemsnitlige scorer for mænd og kvinder for alle årene. Betragtes data fra alle fire år under et er denne forskel på  $6,3 \pm 0,8$  spørgsmål svarende til  $21,0 \pm 2,8\%$ . En sammenligning med adgangsgivende kvotienter viser, at kvindernes lavere gennemsnitlige score ikke lader sig forklare ud fra lavere gennemsnit på de gymnasiale uddannelser, ligesom en sammenligning med eksamenskarakterer fra det introducerende kursus i klassisk mekanik på NBI viser, at kønsforskellen ikke genfindes her. Denne uoverensstemmelse, som også er observeret i en amerikansk kontekst (Docktor & Heller, 2008), er grundlæggende overraskende og gør, at det er væsentligt at forsøge at blive klogere på mulige forklaringsmodeller. De studerendes udeladte svar i FCI-testen er blevet analyseret ved at inddele dem i to kategorier, således at der blev skelnet mellem blanke svar som udtryk for at være løbet tør for tid og som udtryk for en svarstrategi om hellere at ville undlade at svare end at gætte. Det viser sig, at der er flere kvindelige studerende end mandlige der udelader svar. Det viser sig dog også, at denne forskel ikke kan forklare hele den observerede forskel i FCI-resultater. Kvinderne undlod gennemsnitligt at besvare  $0,5 \pm 0,5$  spørgsmål mere end mændene. Dette betyder, at vi med stor sandsynlighed har afvist denne forklaringsmodel og således har brug for nærmere undersøgelser af kønsforskellen i FCI.

Det andet forskningsspørgsmål er besvaret ved hjælp af en analyse af kvalitativ data i form af sociokulturelt orienterede samtaler gennemført med 12 nystartede fysikstuderende. Indsamlingen af denne empiri med baggrund i metodiske overvejelser, der skulle gøre det muligt at mediere to markant forskellige evalueringsparadigmer med tilhørende forskelle i traditioner, videnskabssyn og epistemologisk grundsyn, er derfor i egen ret et væsentligt resultat af dette speciale. Empirimængden er forholdsvis stor og indsamlet i en dansk kontekst, hvilket muliggør, at den senere hen kan bruges til at blive klogere på danske fysikstuderendes forståelse af Newtonsk mekanik i rammer der rækker ud over de FCI-relaterede forskningsspørgsmål i dette speciale. Empiri af en sådan karakter har ikke tidligere været indsamlet på NBI. Ud fra denne empiri blev det konkluderet, at evalueringsformatet har en væsentlig indflydelse på de studerendes besvarelser af FCI-spørgsmålene, hvilket ud fra et sociokulturelt perspektiv forstås ud fra de studerendes adgang til medierende artefakter og meningsskabende dialog – den kommunikative situation er ændret. Herudover sås det, at

også mere pragmatiske forklaringer som for eksempel bedre tid og mindsket betydning af informantens læsekompetence havde en indflydelse på denne ændring. Dette betyder, at fortolkningen af FCI-data må tages med det forbehold, at billedet af studerendes mekanikforståelse er præget af den metode man vælger at evaluere den med. Brugen af FCI som diagnostisk værktøj til at efterprøve den enkelte studerendes forståelse af mekanik med skal altså tages med forbehold for, at denne forståelse måles i et postpositivistisk evalueringsparadigme med alt hvad et sådant paradigme indebærer af både begrænsninger og muligheder.

Ligeledes blev det konkluderet, at ændringerne af besvarelsenerne var gennemgående hos alle de udvalgte cases. Det er ikke ud fra undersøgelsen muligt at afvise, at en ændring af evalueringsformatet kan have en indflydelse på kønsforskellen i FCI-resultater, men omvendt er der heller ikke gjort observationer der peger på, at dette er tilfældet. For at forstå kønsforskellen er der andre sammenhænge, det er mere frugtbart at søge at afdække. Den gennemgående, ikke kønsspecifikke ændring af resultater betød, at specialets fokus gradvist forskød sig fra at omhandle FCI i et kønsperspektiv til at omhandle forståelsen af interessante evaluerings- og læringsteoretiske aspekter af empirien.

Specialet indeholder således en læringsteoretisk analyse af informanternes besvarelse af spørgsmål 21, som ikke direkte relaterer sig til de opstillede forskningsspørgsmål. Det blev af denne analyse konkluderet, at der er en markant kontekstafhængighed i de studerendes konceptuelle forståelse. Misconceptions, der eksisterer i en bestemt kontekst eksisterer ikke nødvendigvis i en anden, mere genkendelig kontekst, hvilket er vanskeligt forklarligt ud fra FCI-testens læringsteoretiske forståelse af, hvordan de studerendes forkerte besvarelser kan ses som et udtryk for eksistensen af grundlæggende misconceptions. Dette betyder, at specialets empiri peger i retning af, at der er behov for en øget forståelse af, hvordan vi skal forstå FCI-besvarelsenerne som resultat af læringsprocesser hos de studerende.

I kapitel 11.2 sås det, at kognitivt orienterede læringsperspektiver omkring eksemplarer, videnstransfer og p-prims bidrog til forståelsen af informanternes besvarelse af spørgsmål 21.

Opsummerende betyder det altså, at en frugtbar symbiose af mental- og socialkonstruktivistisk læringsteori bidrager med interessante perspektiver på, hvordan FCI-testens resultater skal forstås og hvordan disse kan bruges til at blive klogere på, hvordan danske fysikstuderende på universitetsniveau konstruerer viden om Newtonsk mekanik.

## 15 Litteraturliste

Ben-Shakhar, G. & Sinai, Y.:

"Gender Differences in Multiple-Choice Tests: The Role of Differential Guessing Tendencies", *Journal of Educational Measurement*, **28** (1), 23-35 (1991)

Birch, M. & Walet, N.:

"Gender differences in students' conceptual understanding of Newtonian Mechanics" (2012)

[http://www.heacademy.ac.uk/assets/documents/stem-conference/STEMLearningandTeachingIssues1/Marion\\_Birch.pdf](http://www.heacademy.ac.uk/assets/documents/stem-conference/STEMLearningandTeachingIssues1/Marion_Birch.pdf) (hentet 9/1 2013)

Bryman, Alan:

"Quantitative data analysis" i *Social Research Methods* (329-352)

4. udgave

2012a, Oxford University Press

Bryman, Alan:

"Interviewing in qualitative research" i *Social Research Methods* (468-498)

4. udgave

2012b, Oxford University Press

Danielsson, A. T.:

*Doing Physics – Doing Gender*

2009, Uppsala University Library

Docktor, J. & Heller, K.:

"Gender Differences in Both Force Concept Inventory and Introductory Physics Performance" (2008)

[http://groups.physics.umn.edu/physed/Talks/Docktor\\_Heller\\_PERC08.pdf](http://groups.physics.umn.edu/physed/Talks/Docktor_Heller_PERC08.pdf) (hentet 9/1 2013)

Dolin, J.:

"Læringsteorier" i Damberg, E., Dolin, J. & Ingerslev, G. H. (Red.): *Gymnasiepædagogik. En Grundbog* (140-182)

1. udgave

2006, Hans Reitzels Forlag

Dolin, J. & Krogh, L. B.:

"PISA 2006 science testen og danske elevers naturfaglige formåen. Rapport nummer tre fra Validering af PISA-projektet" (2011)

[http://www.ind.ku.dk/publikationer/inds\\_skriftserie/vap3/](http://www.ind.ku.dk/publikationer/inds_skriftserie/vap3/) (hentet 9/1 2013)

Flyvbjerg, B.:

"Five Misunderstandings About Case-Study Research", *Qualitative Inquiry*, **12**, Nr. 2, 219-245 (2006)

Gipps, C.:

"Chapter 10: Socio-Cultural Aspects of Assessment", *Review of Research in Education*, **24**(1), 355-391 (1999)

Halloun, I. A. & Hestenes, D.:

"The initial knowledge state of college physics students", *American Journal of Physics*, **53**(11), 1043-1055 (1985a)

Halloun, I. A. & Hestenes, D.:

"Common sense concepts about motion", *American Journal of Physics*, **53**(11), 1056-1065 (1985b)

Halloun, I. & Hestenes, D.:

"The search for conceptual coherence in FCI data" (1996)  
<http://modeling.asu.edu/R%26E/CoherFCI.pdf> (hentet 26/12 2012)

Hammer, D.:

"More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research", *American Journal of Physics*, **64**(10), 1316-1325 (1996)

Heller, P. & Huffman, D.:

"Interpreting the Force Concept Inventory. A Reply to Hestenes and Halloun", *The Physics Teacher*, **33** (8), 503, 507-511 (1995b)

Hestenes, D. & Halloun, I.:

"Interpreting the Force Concept Inventory. A response to Huffman and Heller", *The Physics Teacher*, **33** (8), 502-506 (1995)

Hestenes, D., Wells, M. & Swackhamer, G.:

"Force Concept Inventory", *The Physics Teacher*, **30**, 141-158 (1992)

Huffman, D. & Heller, P.:

"What Does The Force Concept Inventory Actually Measure?", *The Physics Teacher*, **33** (3), 138-143 (1995a)

Kost, L. E., Pollock, S. J. & Finkelstein, N. D.:

"Characterizing the gender gap in introductory physics", *Physical Review Special Topics – Physics Education Research*, **5**, 010101-1 – 010102-14 (2009)

Kuhn, T. S.:

*The structure of scientific revolutions*

4. udgave

2012, The University of Chicago Press

- Lave, J. & Wenger, E.:  
*Situated learning. Legitimate peripheral participation*  
1. udgave  
1991, Cambridge University Press
- Leach, J. & Scott, P.:  
"Individual and Sociocultural Views of Learning in Science Education", *Science & Education*, **12**, 91-113 (2003)
- Lorenzo, M., Crouch, C. H. & Mazur, E.:  
"Reducing the gender gap in the physics classroom", *American Journal of Physics*, **74**(2), 118-122 (2006)
- McCullough, L.:  
"Gender, Context and Physics Assessment", *Journal of International Women's Studies*, **5** (4), 20-29 (2004)
- Nielsen, A. M.:  
*Rapport om udvikling af værktøj til at forstå studerendes holdninger til fysik og det at lære fysik*, 2004  
<http://www.ind.ku.dk/udvikling/projekter/projekter-2003og2004/rapport-udv.afvrktj.pdf/>  
(hentet 9/1 2013)
- Pollock, S. J., Finkelstein, N. D. & Kost, L. E.:  
"Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?" *Physical Review Special Topics – Physics Education Research*, **3**, 010107-1-010107-4 (2007)
- Popp, S., Meltzer, D. & Megowan-Romanowicz, C.:  
"Is the Force Concept Inventory Biased?" (2011)  
<http://modeling.asu.edu/R&E/FCI%20DIFanalysisAERA2011.doc> (hentet 9/1 2013)
- Rebello, N. S. & Zollman, D. A.:  
"The Effect of Distractors on Student Performance on the Force Concept Inventory", *American Journal of Physics*, **72** (1), 116-125 (2004)
- Rebsdorf, S. O.:  
"Test af studerendes begrebsforståelse", *Dansk Universitetspædagogisk Tidsskrift*, **1**(1), 26-30 (2006)
- Royer, J. M., Mestre, J. P. & Dufresne, R. J.:  
"Introduction. Framing The Transfer Problem" i Mestre, J. P. (Red.): *Transfer of Learning from a Modern Multidisciplinary Perspective* (vii-xxvi)  
2005, Information Age Publishing

Savinainen, A. & Scott, P.:

"The Force Concept Inventory: a tool for monitoring student learning", *Physics Education*, **37** (1), 45-52 (2002)

Savinainen, A. & Scott, P.:

"Using the Force Concept Inventory to monitor student learning and to plan teaching", *Physics Education*, **37** (1), 53-58 (2002)

Savinainen, A. & Viiri, J.:

"The Force Concept Inventory as a Measure of Students' Conceptual Coherence", *International Journal of Science and Mathematics Education*, **6**, 719-740 (2007)

Stadler, H., Duit, R. & Benke, G.:

"Do boys and girls understand physics differently?", *Physics Education*, **35** (6), 417-422 (2000)

Steinberg, R. N. & Sabella, M. S.:

"Performance on Multiple-Choice Diagnostics and Complementary Exam Problems", *The Physics Teacher*, **35**, 150-155 (1997)

Stewart, J., Griffin, H. & Stewart, G.:

"Context Sensitivity in the Force Concept Inventory", *Physical Review Special Topics – Physics Education Research*, **3**, 010102-1 – 010102-6 (2007)

Säljö, R.:

*Læring i praksis. Et sociokulturelt perspektiv.*

1. udgave

2003, Hans Reitzels Forlag

Taasoobshirazi, G. & Sinatra, G. M.:

"A Structural Equation Model of Conceptual Change in Physics", *Journal of Research in Science Teaching*, **48**(8), 901-918 (2011)

Winsløw, C.:

*Didaktiske Elementer. En indføring i matematikkens og naturfagenes didaktik*

1. udgave

2006, Forlaget Biofolia

## Appendix A

### Force Concept Inventory-testen i dansk oversættelse























## Appendix B

### Tabeller over FCI-spørgsmål

FCI's forståelse af spørgsmålenes konceptuelle indhold kan analyseres ud fra Table I i (Hestenes et al., 1992), der angiver spørgsmålenes placering i de 6 konceptuelle dimensioner, kraftbegrebet er opdelt i, samt Table II (Hestenes et al., 1992), der angiver hvilke misconceptions de forskellige forkerte svar i FCI antyder er til stede. Disse to tabeller ses herunder. Da tabellerne relaterer til 1992-udgaven af FCI og ikke 1995-udgaven som er anvendt i udarbejdelsen af specialet, har de enkelte opgaver andre numre. En sådan tabel med oversættelse af numre ses også herunder.

**Table I. Newtonian Concepts in the Inventory.**

	<b>Inventory Item</b>
<b>0. Kinematics</b>	
Velocity discriminated from position	20E
Acceleration discriminated from velocity	21D
Constant acceleration entails parabolic orbit	23D, 24E
changing speed	25B
Vector addition of velocities	(7E)
<b>I. First Law</b>	
with no force	4B, (6B), 10B
velocity direction constant	26B
speed constant	8A, 27 A
with cancelling forces	18B,28C
<b>2. Second Law</b>	
Impulsive force	(6B), (7E)
Constant force implies constant acceleration	24E, 25B
<b>3. Third Law</b>	
for impulsive forces	2E, IIE
for continuous forces	13A, 14A
<b>4. Superposition Principle</b>	
Vector sum	19B
Cancelling forces	(9D), 18B, 28C
<b>5. Kinds or Force</b>	
5S. Solid contact	
passive	(9D), (12 B,D)
Impulsive	15C
Friction opposes motion	29C
5F. Fluid contact	
Air resistance	22D
buoyant (air pressure)	12D
5G. Gravitation	
	5D, 9D, (12B,D), 17C, I8B, 22D
acceleration independent of weight	1C, 3A
parabolic trajectory	16B, 23D

**Table II. A Taxonomy of Misconceptions Probed by the Inventory. Presence of the misconceptions is suggested by selection of the corresponding Inventory Item.**

	<b>Inventory Item</b>
<b>0. Kinematics</b>	
K1. position-velocity undiscriminated	208,C,D
K2. velocity-acceleration undiscriminated	20A; 21B,C
K3. nonvectorial velocity composition	7C
<b>1. Impetus</b>	
I1. impetus supplied by "hit"	9B,C; 22B,C,E; 29D
I2. loss/recovery of original impetus	4D; 6C,E; 24A; 26A,D,E
I3. impetus dissipation	5A,8,C; 8C; 16C,D; 23E; 27C,E; 29B
I4. gradual/delayed impetus build-up	6D; 8B,D; 24D; 29E
I5. circular impetus	4A,D; 10A
<b>2. Active Force</b>	
AF1. only active agents exert forces	11B; 12B; 13D; 14D; 15A,B; 18D; 22A
AF2. motion implies active force	29A
AF3. no motion implies no force	12E
AF4. velocity proportional to applied force	25A; 28A
AF5. acceleration implies increasing force	17B
AF6. force causes acceleration to terminal velocity	17A; 25D
AF7. active force wears out	25C,E
<b>3. Action/Reaction Pairs</b>	
AR1. greater mass implies greater force	2A,D; 11D; 13B; 14B
AR2. most active agent produces greatest force	13C; 11D; 14C
<b>4. Concatenation of Influences</b>	
CI1. largest force determines motion	18A,E; 19A
CI2. force compromise determines motion	4C, 10D; 16A; 19C,D; 23C; 24C
CI3. last force to act determines motion	6A; 7B; 24B; 26C
<b>5. Other Influences on Motion</b>	
CF. Centrifugal force	4C,D,E; 10C,D,E
Ob. Obstacles exert no force	2C; 9A,B; 12A; 13E; 14E
<b>Resistance</b>	
R1. mass makes things stop	29A,8; 23A,B?
R2. motion when force overcomes resistance	28B,D
R3. resistance opposes force/impetus	28E
<b>Gravity</b>	
G1. air pressure-assisted gravity	9A; 12C; 17E; 18E
G2. gravity intrinsic to mass	5E; 9E; 17D
G3. heavier objects fall faster	1A; 3B,D
G4. gravity increases as objects fall	5B; 17B
G5. gravity acts after impetus wears down	5B; 16D; 23E

1992 Question #	1995 Question #
1	1
2	4
3	2
4	7
5	13
6	8
7	9
8	10
9	11
10	6
11	28
12	29
13	15
14	16
15	-
16	12
17	3
18	17*
19	-
20	19
21	20
22	30*
23	14
24	21
25	22
26	23
27	24
28	-
29	27

- Question removed from 1995 version

\* Answers changed in 1995  
 New questions in 1995 version: 5, 18, 25, 26

## Appendix C

### Fysikfaglig analyse af udvalgte spørgsmål

#### Spørgsmål 13 og 30

Disse to spørgsmål befinder sig begge i den konceptuelle dimension der omhandler typer af kræfter. Der testes for flere forskellige misconceptions, men gennemgående handler mange af dem om, hvorvidt den studerende har en impetusforståelse, altså en indre tendens til bevægelse der overføres til et objekt når det for eksempel kastes (Halloun & Hestenes, 1985b). Spørgsmålene kan altså kun besvares korrekt hvis den studerende er i stand til at afgøre, at der ikke kan virke en kraft på bolden efter at den ikke længere er i kontakt med henholdsvis hånden der kaster den og ketsjeren der slår til den.

#### Spørgsmål 21, 22, 23 og 24

Disse 4 spørgsmål omhandler forskellige konceptuelle dimensioner, men behandles her under et, da de kontekstuel er sammenhængende og alle behandler det samme rumskib der bevæger sig i det ydre rum.

Spørgsmål 21 placeres i to forskellige konceptuelle dimensioner. For at svare korrekt skal man dels kunne afgøre, at en konstant kraft giver anledning til en konstant acceleration og dels kunne afgøre, at en konstant acceleration giver anledning til en parabelbane.

Spørgsmål 22 placeres i de samme to dimensioner. For at svare korrekt skal man dels kunne afgøre, at en konstant kraft giver anledning til en konstant acceleration og dels kunne afgøre, at en konstant acceleration giver anledning til en ændring i fart.

Spørgsmål 23 og 24 placeres begge i den dimension der handler om Newtons 1. lov hvor ingen kræfter er til stede. I spørgsmål 23 skal man for at svare korrekt kunne afgøre, at dette betyder, at hastighedens retning er konstant, mens man i spørgsmål 24 skal kunne afgøre, at det medfører, at hastighedens størrelse er konstant.

Der kan ligge en del forskellige misconceptions til grund for disse. En del af dem relaterer igen til impetusforståelser. Herudover er der to andre misconceptions der bør nævnes, nemlig i relation til svarmulighed C i spørgsmål 21 at bevægelsen bestemmes af et kompromis mellem kræfterne og i relation til svarmulighed D i spørgsmål 22 at kræften vil forårsage en acceleration indtil der opnås en sluthastighed.

#### Udarbejdelsen af samtaleskemaet

I forhold til at afdække informanternes forståelse af fysikken som spørgsmålene søger at teste for, blev fysikken i spørgsmålene underinddelt i et finere net end inddelingen i FCI tilsiger.

Denne underinddeling kan ses som at forsøg på at indfange fysisk forståelse omkring situationerne der kan opleves som relevant og brugbar for at forstå den fysiske situation, men som der ikke testes for i FCI.

#### Spørgsmål 13

I et Newtonsk verdensbillede er den kinematiske forståelse af kastet tæt knyttet sammen med den dynamiske forståelse af kræfterne på kuglen. I FCI-testen undersøges de studerendes kinematiske forståelse af kastet ikke. I gentestningen blev de derfor bedt om at beskrive hastighed og acceleration undervejs i kastet for at afdække om den kinematiske forståelse er på plads selvom den dynamiske forståelse ikke er det. I relation til dette blev der spurgt til de

studerendes forståelse af Newtons 2. lov som en måde at undersøge på, hvordan de studerende er i stand til at oversætte (eller ikke at oversætte) mellem den kinematiske og dynamiske dimension.

I relation til FCI's indplacering af spørgsmålet i en konceptuel dimension omhandlende kræfter blev det planlagt, at der på forskellige måder skulle spørges ind til informantens forståelse af, hvilke kræfter der virker på kuglen undervejs i kastet.

I den oprindelige FCI-opgave anføres det i opgaveteksten, at der ses bort fra luftmodstanden undervejs i kastet. Der spørges ind til de studerendes forståelse af rimeligheden i dette. Derudover spørges der til, hvilke kræfter der ellers virker på bolden. Dette giver mulighed for at undersøge de misconceptions der optræder i FCI's kategorisering. Angivelsen af, at der virker en kraft fra hånden på bolden antyder en impetusforståelse hos den studerende. Der spørges efterfølgende ind til, hvornår i kastet denne kraft virker og om den er konstant. Desuden spørges der ind til gravitationskraften for at afgøre, hvorvidt den studerende mener, at den er konstant undervejs i kastet.

Således afdækker disse spørgsmål samme konceptuelle temaer som FCI-svarmulighederne, men giver samtidig mulighed for at sætte den studerendes viden ind i en bredere kontekst – hvordan er deres kinematiske forståelse af den fysiske situation?

### Spørgsmål 30

Samtalen omkring dette spørgsmål bygger videre på samtalen omkring spørgsmål 30. Der spørges således direkte til, hvilke kræfter der virker på tennisbolden efter denne er skudt af sted. Dette giver igen mulighed for at afklare impetusforståelsen der indikeres ved angivelsen af, at der eksisterer en kraft fra ketsjeren på bolden. Desuden giver det mulighed for at afklare, hvilke betingelser de studerende mener skal være til stede for at der virker en kraft forårsaget af luften på bolden.

### Spørgsmål 21-24

Samtalen starter her med at søge at afdække, om det grundlæggende præmis, at rumskibet ikke er påvirket af nogen kræfter og derfor bevæger sig med konstant hastighed, er rimelig for de studerende. Herefter bliver informanten bedt om at bevæge rumskibet, som det vil bevæge efter motoren tændes. Dette danner udgangspunkt for en diskussion af, om det for informanten er meningsfuldt at dele bevægelsen op i to komponenter og at anvende Newtons love til at bestemme hastigheder og accelerationer i hver af disse retninger. Denne opdeling og forståelse af, at bevægelsen i de to retninger er uafhængige, er essentiel for at kunne analysere situationen fyldestgørende. Hvis informanten ikke svarer korrekt er der i samtalskemaet to muligheder for at få uddybet informantens forståelse. For det første kan der spørges ind til, hvordan bevægelsen ville være, hvis rumskibet havde været i hvile da motoren blev tændt. Ved således at reducere problemet fra at være todimensionelt til i stedet at være endimensionelt kan informantens kinematiske forståelse af bevægelsens ene komponent afsøges. Desuden kan informanten blive præsenteret for det fysisk ækvivalente spørgsmål 12 for at se, om forståelsen er til stede i et spørgsmål hvis kontekst er væsentlig anderledes, men hvis indhold er ækvivalent – ligesom det bliver muligt at se, om præsentationen af dette eksempel ændrer deres tankegang omkring det oprindelige spørgsmål.

Disse overvejelser omkring spørgsmålenes fysiskfaglige indhold og opbygningen af indholdet af gentestningen mundede ud i et skema, der beskriver interviewets forløb. Dette skema findes i Appendix D sammen med det pointgivningsskema, det er udviklet sideløbende med.

## Appendix D

I dette appendix ses samtaleskemaet i den udformning som det er benyttet under interviewene. Da jeg har udformet samtaleskemaet til eget brug kan dele af det forekomme indforstået. Jeg har dog med vilje undladt at rette i det efterfølgende. Desuden ses et pointgivningsskema. Dette er ikke benyttet i analysen i specialet, men da samtale- og pointgivningsskemaet er udarbejdet sideløbende er begge dele vedlagt.

### Samtaleskema

#### *Indledende:*

Fulde navn, fødselsdato.

Hvornår tog du adgangsgivende eksamen? Hvilken slags?

Har forældre videregående uddannelse?

Hvad skal der ske? Opfordring til at lege med tingene!

Opfordring til at tænke højt, forklare grundigt.

#### *Spørgsmål 13:*

Dreng kaster bold lodret op. Vi kan se bort fra luftmodstanden. Virker det rimeligt?

Beskriv boldens bevægelse. Brug gerne bolden.

Hvad sker med hastigheden undervejs?

Hvilke kræfter virker på bolden?

Vælg det korrekte udsagn ud af 5 mulige. Tænk gerne højt.

Hvorfor vælger du det?

Hvis forkert:

Hvor kommer den opadrettede kraft fra? Er den konstant undervejs?

Hvis korrekt:

Hvordan kan det være at bolden bevæger sig op når der kun virker en kraft ned?

Hvad med kraft fra drengens hånd? Virker den ikke på bolden?

Hvordan er accelerationen undervejs? Hvorfor (Newtons 2. lov)

#### **Koncepter:**

**Konstant kraft giver acceleration.**

**Acceleration er hastighedsændring.**

**Ingen impetus. Kan ikke have kontaktkraft efter bolden er sluppet.**



## **Tyngdekraft er konstant.**

### *Spørgsmål 30:*

Dame spiller tennis i strid blæst. Det lykkes hende alligevel at skyde bolden over nettet.

Hvilke kræfter virker på bolden fra den har forladt ketsjeren til den rammer jorden?

Hvilke kan du selv nævne?

Liste udleveres. Spørg ind.

Hvis ketsjerslaget nævnes:

Er det en konstant kraft? Hvorfor?

Hvis det ikke nævnes:

Hvad med kraft fra ketsjeren? Hvorfor påvirker det ikke?

Har du andre kommentarer til de her to spørgsmål? Andre ting du kommer til at tænke på?

### **Koncepter:**

**Kontaktkræfter kan ikke overføres.**

**Uddybning af impetusforståelsen fra 1. spørgsmål**

### *Spørgsmål 21-24:*

Rumskib i det ydre rum. Ikke påvirket af nogen ydre kræfter.

Bevæger sig fra P til Q. Demonstrer med artefakt.

Hvordan kan det lade sig gøre at rumskibet bevæger sig selvom der ikke er nogen ydre kræfter?

Når rumskibet når Q tændes der for motoren. Rumskibet bliver da påvirket af en kraft.

Vis med artefakt hvordan rumskibet bevæger sig. Tænk højt, forklar hvorfor!

Vælg en af kurverne.

Hvad hvis rumskibet havde været i hvile inden motoren blev tændt?

Kan vi dele bevægelsen op i to retninger?

Hvad sker der med hastigheden i de to retninger?

Hvis vedkommende er helt blank: Hvad sker der med en kanonkugle der affyres? Hvorfor?

Har du hørt om et skråt kast?

Hvad sker med rumskibets fart efter motoren tændes?

Valgmuligheder.

Nu slukker vi igen for motoren i punktet R.

Hvordan vil rumskibet bevæge sig bagefter. Hvorfor?

Valgmuligheder.

Hvis der svares korrekt: Hvorfor ikke 1? Det var jo banen før da der heller ikke virkede nogen kræfter.

Hvis der svares galt: Prøv at tegne hastighedsvektorer på bevægelseskurven.

Hvilken fart vil det have? Fart sammenlignet med farten inden.

Er der andet du brænder inde med?

**Koncepter:**

**Newtons første lov med ingen kræfter.**

**Konstant kraft giver konstant acceleration. Newtons 2. lov.**

**Sammensat bevægelse – skråt kast, parabelbaner.**

**Retning af bevægelse kan aflæses ud fra en bevægelseskurve.**

*Afsluttende:*

Kan du huske nogen af spørgsmålene?

Har du snakket med andre om dem siden testen?

## Pointgivningsskema

### Spørgsmål 13:

- a)  
Forståelse af, hvorfor det er rimeligt at se bort fra luftmodstanden
- b)  
Korrekt beskrivelse af bevægelsen og hastighederne undervejs, herunder at hastigheden er 0 i toppunktet
- c)  
Tyngdekraften i spørgsmålet er (stort set) konstant
- d)  
En konstant kraft giver anledning til en konstant acceleration
- e)  
Kraften fra hånden er en kontaktkraft og påvirker ikke bolden efter den er sluppet

### Spørgsmål 30:

- a)  
Blæsevejret gør, at luftmodstanden ikke kan negligeres
- b)  
Kontaktkraften fra ketsjeren virker ikke efter bolden har sluppet ketsjeren
- c)  
Tyngdekraften virker

### Spørgsmål 21-24:

- a)  
Newtons 1. lov anvendt til at forklare, at raketten kan bevæge sig uden der virker ydre kræfter
- b)  
Bevægelsen efter motoren tændes kan deles i to komponenter
- c)  
Newtons 2. lov anvendt på den vinkelrette retning (evt. formuleret som tilsvarende spørgsmål hvis raketten ligger i hvile)
- d)  
Konstant hastighed og konstant acceleration giver en parabelbane (ordet parabel behøver ikke at optræde)
- e)  
Eventuelt: En affyret kanonkugle følger en parabelbane

f)

Newtons 2. lov anvendt til at finde accelerationen mellem B og C  
Konstant acceleration giver kontinuert voksende fart

g)

Newtons 1. lov anvendt til at afgøre, at bevægelsen fra C og frem er en ret linje

h)

Hastighedens retning fra C og frem er bestemt ved retningen i C

i)

Newtons 2. lov anvendt til at finde accelerationen fra C  
Acceleration lig nul giver konstant fart