

**BAGGRUNDSNOTAT: UDVIKLING OG VALIDERING AF
SURVEYSKALAER TIL IMPACTMÅLING AF TURBOVÆKST**

2024

INDHOLDSFORTEGNELSE

1. INTRODUKTION	3
2. SELF-EFFICACY.....	4
2.1 Teoretisk forståelse af self-efficacy	4
2.2 Self-efficacy målt i en skolekontekst.....	5
2.3 Self-efficacy målt i naturfag	6
2.4 LIFE's måling af science self-efficacy.....	7
3. INTERESSE OG MOTIVATION.....	11
3.1 Teoretisk forståelse af motivation i Self-determination theory	11
3.2 Teoretisk forståelse af interesse ved Fire-fase-modellen	13
3.3 LIFE's teoretiske grundlag for måling af interesse og motivation	14
3.4 LIFE's måling af interesse og motivation	15
4. STATISTISK VALIDERING	20
4.1 Rasch-modellen.....	20
4.2 Datagrundlag	21
4.3 Resultater af valideringsanalyser	22
4.4 Dokumentation af de enkelte skalaer	23
5. SURVEYSKALAER	32
6. LITTERATUR	33
Bilag A1. Proces for udvikling af skalaer	37

1. INTRODUKTION

Rambøll har på vegne af LIFE Fonden gennemført en impactmåling af undervisningsforløbet Turbovækst. Formålet med impactmålingen er blandt andet at måle effekten af forløbet i forhold til LIFE's missionsmål om at "øge børns og unges interesse og motivation for at fordybe sig i naturvidenskab". Derudover har impactmålingen til formål at måle effekten af at deltage i Turbovækst på elevernes udvikling af 'self-efficacy' indenfor naturfagene. Self-efficacy er et centralt effektmål både i LIFE's generiske forandringsteori (som gælder alle forløb) og i den specifikke forandringsteori for Turbovækst.

I effektevalueringer af undervisningsindsatser generelt og af Turbovækst specifikt er det vigtigt at anvende et solidt og teoretisk velbegrunderet måleinstrument for de udfaldsmål, man evaluerer. Derfor har LIFE's impact- og forskningsafdeling udarbejdet dette notat om, hvordan self-efficacy og interesse i og motivation for at fordybe sig i naturvidenskab kan forstås teoretisk og måles. Formålet med notatet er ud fra en teoretisk redegørelse at argumentere for de valg, der ligger til grund for udviklingen af de måleinstrumenter, som LIFE anvender i impactmålingen af Turbovækst.

Notatet har tre hovedkapitler. De to første handler om self-efficacy (kapitel 2) og om interesse og motivation (kapitel 3). I begge kapitler redegør vi først for den teoretiske forståelse for begreberne, og for hvordan de tidligere er blevet målt i en undervisningskontekst. Dernæst argumenterer vi for, hvordan LIFE forstår og måler de tre begreber i impactmålingen af Turbovækst.

I det tredje hovedkapitel (kapitel 4) præsenterer vi resultaterne af den statistiske validering, LIFE har foretaget af surveyskalaerne. Kapitlet indleder med en beskrivelse af Rasch-modellen, som vi har benyttet som udgangspunkt for skalavalidering. Dernæst følger en beskrivelse af det datagrundlag, skalavalideringen er foretaget på. Endelig præsenterer vi resultaterne af valideringsanalyserne; først de samlede resultater, derefter for de enkelte skalaer.

Sidst i notatet ses samlet de endelige surveyskalaer, LIFE har brugt i impactmålingen af Turbovækst.

2. SELF-EFFICACY

Der er i den senere tids uddannelsesforskning kommet et større fokus på betydningen af elevers non-kognitive færdigheder (fx selvkontrol, motivation, tro på egne evner, etc.) for deres præstationer og valg i uddannelsessystemet. Forskning peger på, at non-kognitive færdigheder har betydelige langtidseffekter på uddannelsesoutcomes (se fx Heckman & Krautz 2012; Makransky et al. 2020). Troen på egne evner, også kaldet *self-efficacy*, er en af de non-kognitive færdigheder, som har vist sig at være god til at forudsige præstationer og adfærd i uddannelsessystemet (se fx Marsh et al. 2018; Richardson et al. 2012). Forskning har desuden vist, at elevers tro på egne evner inden for naturfagene specifikt, ofte kaldet *science self-efficacy*, har stor betydning for deres præstationer i disse fag, herunder hvorvidt de lykkes med at gennemføre naturfagsaktiviteter og -opgaver (se fx Honicke & Broadbent 2016; Ferla, Valcke, & Cai 2009; Tuan et al., 2005; Baldwin et al., 1999; Britner & Pajares, 2006).

Der er to primære årsager til, at *science self-efficacy* er relevant at bruge som outcomemål til en effektmåling af undervisningsforløbet Turbovækst. For det første er det et klart mål, at *science self-efficacy* blandt eleverne øges af at deltage i Turbovækst. *Self-efficacy* er et centralt effektmål både i LIFE's generiske forandringsteori (som gælder alle forløb) og i den specifikke forandringsteori for Turbovækst. For det andet viser forskning, at elevers akademiske *self-efficacy* har en væsentlig betydning for elevernes senere præstationer (se fx Marsh et al. 2018; Richardson et al. 2012; Honicke & Broadbent 2015). Zoomer man ind på det forskningsfelt, der mere specifikt undersøger *self-efficacy* inden for naturfag, er der ligeledes evidens for, at *science self-efficacy* har betydning for præstationer i naturfag (Ferla et al. 2009; Hu et al. 2022; Britner & Pajares, 2006). Desuden viser forskningen, at elevers *self-efficacy* på kort sigt kan forandres gennem undervisning (Hu et al. 2022). *Self-efficacy* er derfor velegnet som outcomemål for indsatser som Turbovækst, der gennemføres over en relativt kort periode.

I dette kapitel udfolder vi *self-efficacy*-begrebet samt de meget forskellige måder det tidligere i forskningen er målt på. Sidst i kapitlet beskriver vi, hvordan LIFE ved hjælp af tre surveyskalaer måler *science self-efficacy* i impactmålingen af Turbovækst. Bilag A1 beskriver de trin og metoder der er anvendt i udviklingen af skalaerne.

2.1 Teoretisk forståelse af *self-efficacy*

Vi har i LIFE gennemført et litteraturstudie ved systematisk at søge efter fagfællebedømte forskningsartikler, der måler elevers *self-efficacy* inden for STEM-fagene (se bilag A1 for uddybning). Langt hovedparten af forskningsartiklerne henviser til Albert Banduras forståelse af *self-efficacy* som udgangspunkt for deres anvendelse af begrebet. Der er således teoretisk konsensus om, at Banduras forståelse af begrebet i høj grad udgør den teoretiske ramme om den senere forsknings måling af begrebet. De identificerede artikler henviser primært til Banduras bog "Self-efficacy – the exercise of control" (1997), som er citeret over 120.000 gange. I det følgende diskuterer vi Banduras forståelse af *self-efficacy*.

Bandura definerer *self-efficacy* som følger:

"Perceived self-efficacy is concerned with people's beliefs in their capabilities to produce given attainments."
(Bandura 1997).

Ifølge Bandura handler *self-efficacy* om tro på egne evner til at opnå bestemte resultater. En persons *self-efficacy* vil variere på tværs af livets mange domæner, da man ikke kan udvikle evner inden for alle livets domæner. Ifølge Bandura er *self-efficacy* derfor ikke en overordnet egenskab, men skal ses som et differentieret sæt af overbevisninger om egne evner, der kobler sig til bestemte funktions- eller aktivitetsdomæner. Fx nævner Bandura, at en erhvervsleder kan have en høj *self-efficacy* i at lede en virksomhed men en lav *self-efficacy* i at være forælder.

Bandura lægger altså vægt på, at self-efficacy skal forstås domænespecifikt. Samtidig lægger han vægt på, at begrebet ikke må reduceres til en tro på meget afgrænsede subfærdigheder. Self-efficacy er ikke et mål for, hvilke konkrete færdigheder personen tror han/hun har men snarere for, hvad personen tror han/hun kan gøre under forskellige omstændigheder med de færdigheder, personen besidder. Fx nævner Bandura, at hvis man skal måle en persons tro på egne evner til at køre bil, skal man ikke spørge om personen kan tænde bilen, dreje rattet, skifte gear og aflæse vejskilte. I stedet skal man spørge, om personen kan navigere bilen effektivt under forskellige trafikale omstændigheder med forskellige sværhedsniveauer. Mange forskere begår ifølge Bandura den fejl at reducere self-efficacy til et mål for separate subfærdigheder (Bandura 1997).

Der er altså en balance i at forstå self-efficacy som noget, der knytter sig til et bestemt aktivitetsdomæne, uden derved at reducere det til troen på at kunne udføre afkoblede, specifikke aktiviteter.

2.2 Self-efficacy målt i en skolekontekst

Når man skal udvikle et survey-mål, der måler self-efficacy, er der to overordnede spørgsmål, som er væsentlige at forholde sig til. For det første må man forholde sig til, hvor *specifikt* målet for self-efficacy skal være. Skal det måle en generel skole self-efficacy, self-efficacy inden for naturfag eller bestemte kompetencer inden for naturfag? For det andet er det vigtigt at afgøre, hvad der skal karakterisere udformningen af de konkrete items i et survey-mål, der måler self-efficacy.

Disse to overordnede spørgsmål har været genstand for megen diskussion i self-efficacy-forskningen (se fx Bong og Skaalvik 2003; Marsh et al. 2018). Til at besvare spørgsmålene vil vi fremlægge Banduras perspektiver, men vi vil også inddrage forskningsartikler, der specifikt beskæftiger sig med self-efficacy målt i en skolesammenhæng.

Tre niveauer af mål for self-efficacy

Ifølge Bandura skal målet for self-efficacy bindes op på et bestemt funktionsdomæne (Bandura 1997). Indenfor dette domæne kan det imidlertid variere, hvor generelt eller specifikt et self-efficacy-mål, der er behov for. Han sonderer mellem tre niveauer:

1. **Specifikt mål for self-efficacy:** måler personens tro på egne evner i forhold til *en bestemt* præstation under nogle specifikke omstændigheder.
2. **Mellemniveau for mål af self-efficacy:** Måler personens tro på egne evner i forhold til *en gruppe af* præstationer inden for samme aktivitetsdomæne, der gennemføres under nogle bestemte forhold.
3. **Generelt mål for self-efficacy:** Måler self-efficacy uden at specificere aktiviteter eller omstændigheder.

Hvilket niveau man vælger at måle self-efficacy på afhænger af, hvad man ønsker at forudsige: Jo mere specifikke præstationer man ønsker at forudsige, jo mere specifikke mål. På trods af denne åbning for, at self-efficacy kan måles på forskellige niveauer, advarer Bandura som tidligere nævnt mod at anvende så generelt et self-efficacy-mål, at det ikke kan knyttes an til et afgrænset aktivitetsdomæne.

Når det specifikt kommer til elevers self-efficacy i skolen, advarer Bandura mod at bruge et generelt mål for elevens self-efficacy på tværs af fag, da self-efficacy i skolen kan variere betydeligt fra fag til fag. Hvis man fx skal forudsige præstationer og faglig udvikling i faget matematik, er det vigtigt at der udvikles et mål for self-efficacy, der er rettet mod de kompetencer, der skal anvendes i matematik (Bandura 1997). I forbindelse med LIFE's effektmåling af Turbovækst er det altså vigtigt at målet for self-efficacy er rettet mod naturfag.

Self-efficacy overfor self-concept

Det at måle elevers self-efficacy i skolen beskrives ofte i forskningen som akademisk self-efficacy (Honicke & Broadbent 2015). Inden for denne forskning er der fokus på at identificere det særegne ved begrebet akademisk self-efficacy i forhold til relaterede begreber som akademisk *self-concept* (akademisk selvopfattelse) (Marsh et al. 2018;

Bong & Skaalvik 2003; Ferla et al. 2009). I Marsh et al.'s bog "Handbook of Competence and Motivation" (2018) udfoldes to nøglekarakteristika ved akademisk self-efficacy som adskiller det fra akademisk selvopfattelse.

For det *første* er akademisk self-efficacy prospektivt, dvs. det angår den enkeltes opfattelse af, hvad vedkommende - indenfor specifikke opgaver og undervisningsaktiviteter og i specifikke kontekster -- kan opnå aktuelt eller i nær fremtid. Items bør derfor formuleres som spørgsmål om, hvad respondenteren "kan" i relation til en specifik undervisningsopgave/aktivitet. For det *andet* skal items konstrueres, så svaret ikke afhænger af en sammenligning med andres evner. Det skal altså være muligt at vurdere, om man kan klare en opgave/aktivitet, uden at svaret påvirkes af sammenligning med andres præstationer. En del mål for akademisk self-efficacy indeholder items, der afføder social sammenligning som fx "Jeg er sikker på, at jeg klarer mig virkelig godt i opgaver og tests". En elevs vurdering af, hvorvidt han/hun klarer sig "virkelig godt" vil med stor sandsynlighed afhænge af, hvor dygtige de andre elever i klassen er (Marsh et al. 2018). Desuden vil svaret formentlig afhænge af elevens forventninger til sig selv. Den type betragtninger hører ikke under akademisk self-efficacy men snarere akademisk selvopfattelse (Marsh et al. 2018; Bandura 1986). Items for akademisk self-efficacy bør altså kunne besvares ved at vurdere ens egne evner til at klare en undervisningsaktivitet/opgave uden at vurderingen relateres til andres evner eller egne standarder.

Akademisk selvopfattelse bør omvendt måles med items, der retter sig mod en mere generel opfattelse af egne kompetencer i et fag eller fagområde (og ikke specifikke undervisningsopgaver eller -aktiviteter) samt items, til hvilke svaret ofte vurderes ved sammenligning med andres kompetencer (Marsh et al. 2018; Bong & Skaalvik 2003; Schunk & Pajares 2002). I forskningen er der indikationer på, at de to begreber ikke kun adskiller sig teoretisk men også empirisk. Det lader yderligere til, at self-efficacy er bedst til at forudsige akademiske præstationer, mens akademisk selvopfattelse er bedst til at forudsige affektive faktorer som motivation (Ferla et al. 2009).

2.3 Self-efficacy målt i naturfag

Ovenfor har vi fremhævet enkelte centrale overvejelser, man bør gøre sig i forbindelse med en måling af self-efficacy i en skolekontekst. I det følgende afsnit zoomer vi ind på, hvordan self-efficacy tidligere er blevet målt inden for de naturvidenskabelige fag (ofte kaldet *science self-efficacy*). For at undersøge dette har vi gennemført et litteraturstudie, hvor vi har søgt efter fagfællebedømte forskningsartikler, der har målt eleveres self-efficacy specifikt inden for naturfag.

Vi har identificeret en række forskningsartikler, hvori science self-efficacy måles med en surveyskala. Af vores gennemlæsning af artiklerne står det klart, at self-efficacy måles på mange forskellige måder. Helt overordnet varierer måleinstrumenterne på to parametre. For det første er der variation i, hvor specifikt målet for self-efficacy er. Nogle måleinstrumenter måler bredt set self-efficacy inden for de naturvidenskabelige fag (se fx Lamb et al. 2014), mens andre måler self-efficacy inden for specifikke fag som fysik eller kemi (se fx Tezer & Asiksoy 2015; Baldwin et al. 1999), og endnu andre måler specifikke kompetencer, der knytter sig til specifikke indsatser (Fx bygge robotter; se Tsai et al. 2021). For det andet er der variation i antallet af dimensioner, science self-efficacy måles på. Nogle måler science self-efficacy som et endimensionelt begreb (se fx Lamb et al. 2014), mens andre måler det som flerdimensionelt. Fx måler Baldwin et al. (1999) science self-efficacy indenfor biologi på tre dimensioner: Metoder, generalisering til andre biologi- eller science-fag, og applicering af biologiske begreber og kompetencer.

Der er altså forskelle i, hvor specifikt et mål for self-efficacy der arbejdes med, og hvor mange underdimensioner, det består af. Hvor man placerer sig på disse spørgsmål afgøres i høj grad af formålet med forskningsprojektet.

På tværs af disse forskelle er det tydeligt, at de enkelte items i måleinstrumenterne varierer alt efter, hvorvidt målingen er rettet mod specifikke opgaver/aktiviteter i naturfagsundervisningen eller mod en mere generel tro på egne evner i et naturfag. I den ene yderlighed opstilles items som konkrete opgaver (fx en opgave i brøkgregning), som eleven bedes vurdere sin tiltro til at kunne løse. I den anden ende af kontinuummet spørges der til en generel tiltro til egne præstationer i naturfag (fx Jeg forventer at klare mig rigtigt godt i matematik). Selvom målet for selv-

efficacy er rettet mod samme domæne (fx matematik), er der altså stor forskel på, hvor aktivitetsrettede items i måleinstrumenterne er. Måleinstrumenter, der retter sig mod en generel tro på egne evner i naturfag, snarere end enkelte aktiviteter, er ofte inspireret af måleredskabet Motivated Strategies for Learning Questionnaire (MSLQ) udviklet af Pintrich et al. (1991). Man kan argumentere for at denne type måleinstrumenter i højere grad måler akademisk selvopfattelse og ikke self-efficacy (se ovenstående afsnit).

Udover denne forskel lader der til at være endnu en væsentlig forskel mellem måleinstrumenter, der måler science self-efficacy. Nogle måleinstrumenter fokuserer på at måle self-efficacy inden for specifikke *faglige emner* (fx 'jeg kan forklare hvad lys er'), mens andre fokuserer på kompetencer, der går *på tværs af faglige emner* (fx 'jeg kan opstille hypoteser om udfaldet af en undersøgelse'). Denne forskel er i mindre grad genstand for diskussion i forskningslitteraturen.

Helt overordnet kan måleinstrumenterne placeres i en af følgende tre tilgange:

1. **Generel tiltro til egne evner i naturfag/bestemte naturfag.** Måler en generel tiltro til at kunne opnå en præstation i et eller flere naturfag. Måleinstrumenter i denne kategori er ofte inspireret af måleredskabet MSLQ (Pintrich et al. 1991). Fx *Jeg tror jeg vil få en god karakter i kemi* (se fx Boz et al. 2016; Usher & Pajares 2009; Sadi & Çakiroğlu 2014).
2. **Aktivitetsspecifik måling som ikke er relateret til faglige emner.** Måler tiltro til at kunne udføre undervisningsaktiviteter/opgaver, der ikke er relateret til et fagligt emne. Fx *Konstruere en model til at forklare et naturfænomen* eller *Hvor godt kan du fortolke kemiske formler* (se fx Baldwin et al. 1999; Hu et al. 2022; Lin & Tsai 2013; Tezer & Asiksoy 2015; Uzuntiryaki & Aydin 2009; Tatar et al. 2009).
3. **Aktivitetsspecifik måling som relaterer sig til faglige emner.** Måler tiltro til at kunne udføre undervisningsaktiviteter/opgaver, der direkte relaterer sig til et fagligt emne. Fx *Jeg kan forklare forskellen på varme og temperatur* (se fx Fidan & Tuncel 2021; Summers & Falco 2022; Tsai et al. 2021).

Ovenstående tre tilgange er analytiske kategorier. I praksis vil nogle forskningsartikler anvende spørgsmål fra flere end en tilgang. Referencerne er placeret hvor de passer bedst ind.

I en dansk forskningssammenhæng er det værd at nævne to måleinstrumenter. For det første er der til forskningsprojektet SCOPE udviklet et måleinstrument, der måler *science kapital* (SCICAP). Som en underdimension måles "Selvvurderede færdigheder inden for science" (Keilow et al. 2023). For det andet har Nordic Metrics udviklet 19 non-kognitive færdighedsskalaer: de såkaldte UiL-skalaer. En af disse måler self-efficacy generelt i skolen, og Nordic Metrics har senere udviklet en self-efficacy-skala rettet mod matematik (Makransky et al. 2019). Både Nordic Metrics' og SCOPEs måleinstrumenter er forskningsvaliderede skalaer. Begge skalaer passer bedst i den første af de tre ovenstående tilgange og måler altså primært en generel tiltro til egne evner i naturfag. Nordic Metrics' skala er inspireret af måleredskabet MSLQ (Makransky et al. 2019).

2.4 LIFE's måling af science self-efficacy

Den overordnede ramme for LIFE's mål for science self-efficacy er inspireret af Banduras forståelse af self-efficacy. LIFE forstår således science self-efficacy som elevens tro på egne evner til at opnå resultater inden for aktivitetsdomænet naturfag. LIFE har derfor udviklet et måleinstrument rettet mod præstationer, der kobler sig til aktiviteter og kompetenceområder inden for naturfag.

LIFE er inspireret af de to aktivitetsrettede tilgange til at måle science self-efficacy, som svarer til tilgang 2 og 3 i forrige afsnit. Der er udviklet et måleredskab for hver af de to tilgange. Vi har valgt at bruge en svarskala der minder om den svarskala Baldwin et al. (1999) bruger, og som også minder om den svarskala Bandura forslår at bruge til børn.

Nedenfor gennemgår vi, hvordan de to måleredskaber er udviklet, og hvilke items de består af.

Aktivitetsspecifik måling som ikke er relateret til faglige emner (tilgang 2)

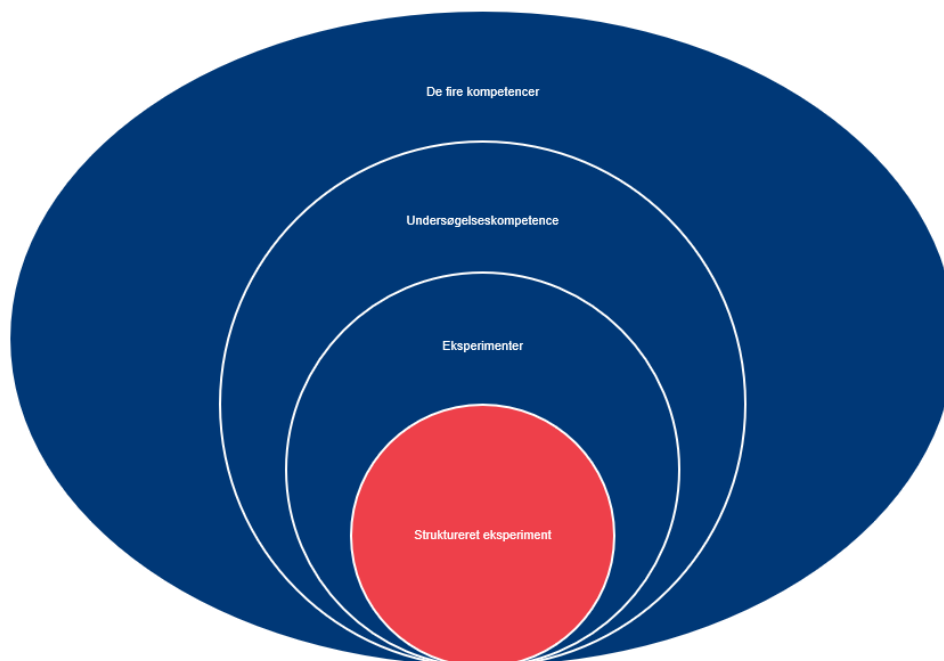
Inden for denne tilgang til måling af self-efficacy spørges ind til aktiviteter, der knytter an til generiske kompetencer, som eleverne skal beherske på tværs af de naturfaglige emner. I den internationale forskning der tager udgangspunkt i denne tilgang, er det forskelligt hvilke generiske kompetencer målingerne retter sig mod. I en dansk skolekontekst er det oplagt at tage udgangspunkt i de fire bindende kompetenceområder (undersøgelse, modellering, perspektivering og kommunikation) og de dertilhørende kompetencemål, som går på tværs af alle tre naturfag (Fælles Mål fysik/kemi, biologi, geografi i 9. klasse, 2019). Læringsmålene for Turbovækst er direkte koblet til de fire kompetenceområder og -mål (jf. lærervejledningen til Turbovækst).

Det er ikke praktisk muligt at måle alle fire kompetenceområder grundigt, når spørgeskemaet også skal indeholde mål for elevernes interesse og motivation for naturfag. LIFE har derfor valgt at fokusere på et af de fire kompetenceområder, nemlig undersøgelseskompetencen. Alle fire kompetenceområder er vigtige i LIFE forløbene, men vi har valgt at fokusere på at måle undersøgelseskompetencen, da LIFE forløbene særligt har fokus på at eleverne gennemfører eksperimenter. Undersøgelseskompetencen er et vidt begreb der i sig selv indeholder mange aspekter (Elmose, 2015). Vi har valgt at fokusere på den del af undersøgelseskompetencen, som handler om at gennemføre eksperimenter i et laboratorium. Det skyldes at det i høj grad er den del af undersøgelseskompetencen, der trænes i Turbovækst.

I Turbovækst Kit gennemfører eleverne tre eksperimenter over fem uger. Dertil kommer, at knap 50% af klasserne, der har gennemført kitforløbet, vælger at supplere det med Turbovækst Lab, hvor de på en dag gennemfører yderligere to eksperimenter. Flere af eksperimenterne er omfattende og rummer flere deleksperimenter. De tre til fem eksperimenter, som eleverne gennemfører i Turbovækst, anses for at være høj i forhold til den ordinære undervisning der gennemføres i naturfag. LIFE har gennemført desk research af de eksperimenter, der gennemføres i Turbovækst, og ud fra Riga et al. (2017) kategoriseret eksperimenterne i følgende fire kategorier: lukket, struktureret, guidet og åben. Undersøgelsen viste, at alle eksperimenter på nær et kan karakteriseres som strukturerede eksperimenter. Se rapporten 'Skolernes undervisning i LIFE's forløb Turbovækst' for en uddybning af forløbets indhold og hvordan det bringes i spil i undervisningen¹. Items i måleinstrumentet er derfor rettet mod at måle elevens tiltro til at gennemføre strukturerede eksperimenter. Måleinstrumentet måler altså en mindre del af elevernes undersøgelseskompetence, hvilket er vist i nedenstående figur.

Figur 2.4.1: Måleinstrumentet måler self-efficacy i forhold til at gennemføre strukturerede eksperimenter (rød cirkel).

¹ 'Skolernes undervisning i LIFE's forløb Turbovækst' kan tilgås på www.life.dk.



Det der kendetegner det strukturerede eksperiment, er at undersøgelsesspørgsmålet og forsøgsopstillingen er givet på forhånd, men at eleverne ikke kender resultatet på forhånd. Eleverne skal selv gennemføre eksperimentet med udgangspunkt i en forsøgsvejledning, og selv reflektere over hvad eksperimentet viser. I en stor del af forsøgene skal eleverne også tage stilling til, hvilken hypotese de vil afprøve i forsøget. For at gennemføre et struktureret eksperiment kræver det blandt andet, at eleverne forstår hvad en hypotese er, kan fortage variabelkontrol, gennemføre relevant dataindsamling samt kan fortolke resultaterne af forsøget. Skalaen fremgår af Tabel 1.

Tabel 2.4.1: Skala til måling af self-efficacy – eksperimenter

Self-efficacy – eksperimenter (5)

Forestil dig, at du skal lave et forsøg i Fysik/kemi.

Vurder, hvor sikker du er på, at du kan:

Lave en hypotese om, hvad du tror, resultatet af forsøget bliver

Fortælle, hvilke variable (fx temperatur) du ændrer i forsøget, og hvilke variable du ikke ændrer i forsøget

Gennemføre dataindsamling/målinger, du kan bruge til at vurdere resultatet af forsøget

Forklare, hvad forsøget viser

Forklare, hvordan forsøget gør dig klogere på det emne, du har om i fysik/kemi

Aktivitetsspecifik måling som er relateret til faglige emner (tilgang 3)

LIFE har også udviklet et self-efficacy måleinstrument, der retter sig mod faglige emner i Turbovækst. Items i denne kategori er udarbejdet, så de bredt set dækker de opstillede læringsmål i forløbet. Der er fire overordnede læringsmål for kitforløbet og to overordnede læringsmål for labforløbet. Læringsmålene i Turbovækst er knyttet til de vejledende færdigheds- og vidensmål i Fælles Mål for de tre naturvidenskabelige fag (jf. læervejledningen til Turbovækst).

Måleinstrumentets items er udviklet, så der spørges ind til elevernes vurdering af, hvorvidt de kan gennemføre aktiviteter, der knytter sig til Turbovæksts overordnede læringsmål. Der er udviklet et måleredskab til kit forløbet med seks spørgsmål og et måleredskab til lab forløbet med fire spørgsmål.

Tabel 2.4.2: Skala til måling af self-efficacy – læringsmål kit og lab

Self-efficacy – læringsmål kit (6)

Vurder, hvor sikker du er på, at du kan:

- Fortælle, hvad et lands befolkningspyramide viser
- Fortælle, hvad en katalysator er
- Lave et forsøg, der viser om et metal er katalysator i en kemisk reaktion
- Lave et forsøg, der viser, om kunstgødning har betydning for planterens vækst
- Give eksempler på både fordele og ulemper ved at bruge kunstgødning
- Give eksempler på, hvad et land kan gøre for at øge fødevarereproduktionen

Self-efficacy – læringsmål lab (4)

Vurder, hvor sikker du er på, at du kan:

- Give et eksempel på, hvad ammoniak kan bruges til
 - Skrive ammoniak som en kemisk formel
 - Nævne mindst to variable, der er vigtige, når et anlæg producerer ammoniak
 - Lave et forsøg, der viser, om temperatur er vigtig for produktion af ammoniak
-

3. INTERESSE OG MOTIVATION

LIFE tager udgangspunkt i to forskellige teorier om interesse og motivation: Self-determination theory og Fire-fase-modellen for interesseudvikling. Begge teorier er bredt anerkendte af motivationsforskere og understøttes af empirisk forskning. De er også ofte anvendt til at forstå udvikling af interesse og motivation inden for det naturfagsdidaktiske felt i Danmark (Petersen, 2017). De har hver deres perspektiv på udvikling af interesse og motivation men også en række fælles berøringspunkter. De to teorier komplementerer hinanden ved at fokusere på forskellige aspekter i forhold til udvikling af interesse og motivation, der er relevante for LIFE's praksis². Derfor er det oplagt at kombinere dem.

I dette kapitel gennemgår vi først Self-determination theory som den præsenteres i litteraturen. Dernæst følger en tilsvarende gennemgang af Fire-fase-modellen for interesseudvikling. Efter den sammenfatter vi kort LIFE's samlede teoretiske forståelse af interesse og motivation og relaterer den til LIFE's forandringsteori. Til sidst beskriver vi, hvordan LIFE omsætter den teoretiske forståelse til konkrete skalaer, der kan måle interesse- og motivationseffekter af Turbovækst.

3.1 Teoretisk forståelse af motivation i Self-determination theory

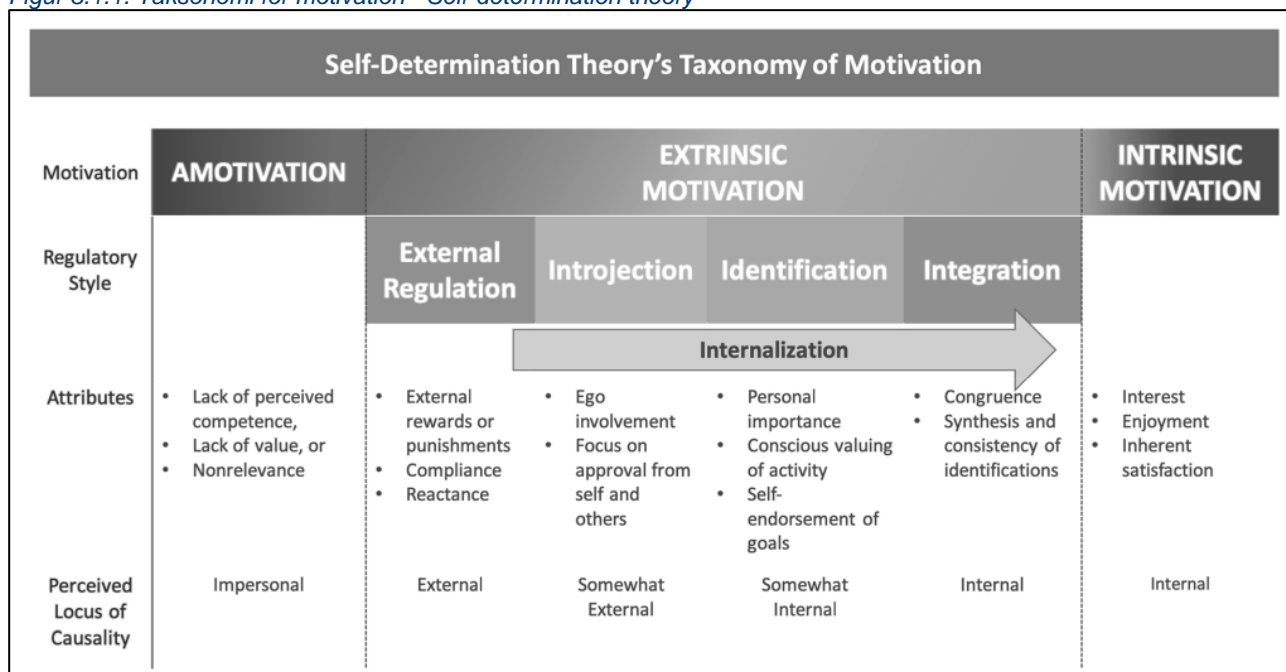
Self-determination theory (Ryan & Deci, 2017; 2020) opstiller en ramme for at forstå, hvilke drivkræfter der ligger bag den enkeltes motivation, dvs. hvad man er motiveret af. Centralt i teorien er ideen om indre motivation, som udgøres af en indefrakommende og iboende drivkraft for at handle og engagere sig, dvs. at gøre noget for aktivitetens egen skyld. Leg, udforskning og handling båret af nysgerrighed er eksempler på aktiviteter drevet af indre motivation, fordi de ikke afhænger af ydre incitamenter eller pres. Dermed er teorien en såkaldt organismeteori, der antager at mennesker har en iboende drift mod at udvikle sig, lære, mestre og forbinde sig til andre. Men disse kræfter udløses ikke automatisk. Det kræver at tre basale psykologiske behov er opfyldt: behovet for autonomi, kompetence og forbundethed.

Autonomi handler om at opleve at have initiativ og ejerskab. I skolesammenhæng understøttes behovet af valgmuligheder og af, at der inden for valgmulighederne er kontekster, som eleven finder interessante og værdifulde. *Kompetence* handler om en følelse af at kunne mestre, vokse og få succes. I skolesammenhæng understøttes behovet bedst af velstrukturerede læringskontekster, der tilbyder optimal udfordring, positiv feedback og muligheder for udvikling. Kompetencebehovet er relateret til Banduras (1997) begreb self-efficacy, der handler om tro på egne evner (jf. kapitel 2). Self-efficacy er et kernebegreb i de fleste motivationsteorier, da følelsen af 'at kunne' øger sandsynligheden for at være motiveret. *Forbundethed* handler om følelsen af at høre til og være forbundet til andre. Det faciliteres af, at de involverede udviser respekt og omsorg også i skolesammenhæng.

I følge Ryan & Deci (2017; 2020) skal alle tre behov være opfyldt, for at elever kan udvikle indre motivation. Er de det, fører det gradvist til en bevægelse fra ydre motivation til indre motivation via en internalisering af opgaveløsningen. Ifølge den bagvedliggende forskning fører indre motivation til bedre præstationer i skolesammenhæng end ydre motivation. Særligt understøttelse af behovet for autonomi er central for udvikling af indre motivation i skolesammenhæng, og aktiviteter drevet af indre motivation hænger ofte sammen med varig læring (Ryan & Deci, 2020).

² Der findes også andre teoretiske skoler inden for interesse og motivationsforskningen, f.eks. Expectancy-Value Theory (Wigfield & Eccles, 2000) og Person-Object Theory (Krapp, 2002). LIFE har reviewet dem, men fundet at de har mindre relevans for LIFE's praksis.

Figur 3.1.1. Taksonomi for motivation - Self-determination theory



Kilde: Ryan & Deci (2020).

I Self-determination theory opstilles en taksonomi for motivation gående fra ikke-motivation i venstre til indre motivation i højre (Figur 3.1.1). I et bredt midterfelt ligger den ydre motivation, der beskriver adfærd drevet af andre incitamenter end at gøre noget udelukkende for aktivitetens egen skyld. Feltet for ydre motivation er bredt, fordi det er heterogent. Dvs. at den ydre motivation varierer i indhold og karakter. Feltet er inddelt i fire undergrupper: *External Regulation*, *Introjection*, *Identification* og *Integration*. De to første undergrupper, *External Regulation* og *Introjection*, udgør den kontrollerede motivation, hvor handlinger udføres for at opnå belønning eller undgå straf. I *Introjection* kan det også handle om at opnå anerkendelse og status. Fælles for de to kontrollerede former for motivation er, at motivationen kommer af at føle sig presset til at udføre en handling enten af andre eller af sig selv. Forskning viser, at denne type af motivation fører til mere overfladisk og mindre overførbare læring end den autonome motivation beskrevet nedenfor (Ryan & Deci, 2017).

De to sidste undergrupper, *Identification* og *Integration*, benævnes autonome, fordi de beskriver en drivkraft, der er initieret af noget ydre – fx det at opnå bestemte karrieremål – men som samtidig i større eller mindre grad opleves som i overensstemmelse med egne værdier eller interesser. *Integration* grænser op til den indre motivation og er karakteriseret af fuld overensstemmelse dvs. integration mellem den ydre motiverede handling og egne værdier og overbevisninger. Det gør, at drivkraften i denne undergruppe er næsten lige så kraftfuld som indre motivation i forhold til at kunne føre til høje præstationer og varig læring (Ryan & Deci, 2020).

Jo længere til højre i taksonomien en motivation kan siges at være reguleret, jo mere har personen internaliseret opgaven og jo mere selvbestemt opleves det at engagere sig i opgaven. Snarere end hvordan motivation udvikles, er taksonomien (Figur 3.1.1) en model for, hvordan forskellige typer af regulering medfører forskellige former for motivation. Dermed ikke sagt, at motivation ikke kan ændre sig i begge retninger. Men ændringer i motivationsform er betinget af ændringer i graden af understøttelse af de tre basale psykologiske behov, som teorien peger på.

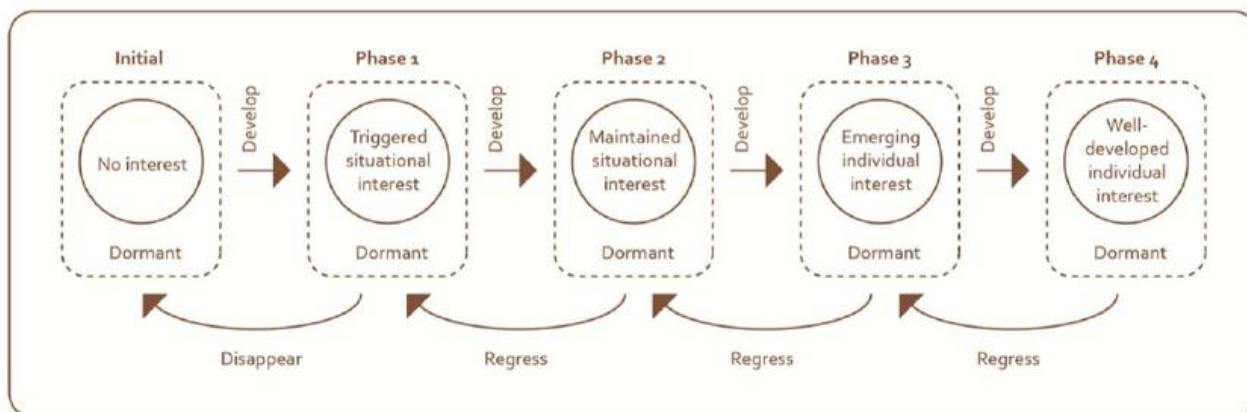
I Self-determination theory er det i skolesammenhæng en kernehypotese, at oplevelsen af selvbestemmelse øger elevens engagement, læring og trivsel. Endvidere at understøttelse af de tre basale psykologiske behov både fra lærere, læringskonteksten og forældre vil facilitere udviklingen af indre motivation, samt at mangel på samme vil føre til ydre- eller ikke-motivation. Hypotesen er understøttet af empirisk forskning på tværs af klassetrin, indholdsområder og kulturelle kontekster (Ryan & Deci, 2020). Fx er det vist, at understøttelse af de tre basale behov er associeret både med større engagement og bedre præstationer indenfor STEM-fag og med større identifikation med det at arbejde inden for STEM (Ryan & Deci, 2020).

LIFE anvender Self-determination theory som sit generelle udgangspunkt for at forstå forskellige former for interesse og motivation i forhold til naturvidenskab. I forhold til hvordan interesse og motivation udvikles hos eleverne, især i relation til et enkelt forløb som Turbovækst, suppleres teorien med fire-fase-modellen for interesseudvikling.

3.2 Teoretisk forståelse af interesse ved Fire-fase-modellen

Fire-fase-modellen for interesseudvikling (herefter Fire-fase-modellen) er funderet i en forståelse af interesse som en psykologisk tilstand af at være interesseret i noget i større eller mindre grad (Hidi & Renninger, 2006). Modellen beskriver udviklingen af interesse i fire faser fra en situationel interesse, der igangsættes af en ydre påvirkning, til en veludviklet individuel interesse, som har udviklet sig til en psykologisk stabil tilstand (Figur 3.2.1).

Figur 3.2.1. Illustration af fire-fase-modellen for interesseudvikling



Kilde: Beh et al. (2015) fortolket ud fra Hidi and Renninger (2006).

Den første fase - *igangsæt situationel interesse* - er en psykologisk tilstand kendetegnet ved følelsen af at være interesseret og generelt at have positive følelser i forhold til et konkret indhold i en given situation. Tilstanden er karakteriseret af flygtige affektive og kognitive påvirkninger udløst af en påvirkning udefra. Det kan være kontekster og aktiviteter, der opleves som uforklarlige, overraskende, personligt relevante, intense og/eller som man kan relatere til. Igangsæt situationel interesse kan forlænges gennem fortsat ydre påvirkning, der opleves som meningsfuld og/eller indebærer involvering og inddragelse. Hvis dette sker, kan tilstanden udvikle sig til *fastholdt situationel interesse*, som er modellens anden fase (Hidi & Renninger, 2006).

Gennem fortsat ydre påvirkning, fx udfordrende og engagerende læringskontekster, kan en fastholdt situationel interesse udvikle sig til den tredje fase i modellen, *emergerende individuel interesse*, som er karakteriseret ved en begyndende nysgerrighed og en prædisposition til at engagere sig i lignende opgaver. Emergerende individuel interesse kan udvikle sig til den fjerde fase i modellen, *veludviklet individuel interesse*, som fortrinsvis genereres af individet selv (Hidi & Renninger, 2006).

Veludviklet individuel interesse er ifølge teorien en relativt stabil psykologisk tilstand karakteriseret af positive følelser og etableret viden om og værdsættelse af et bestemt indholdsområde, fx naturvidenskab. Tilstanden både *fører til* og *næres af* bevidste valg om at engagere sig i bestemte aktiviteter, når lejligheden byder sig, eller helt på egen hånd ud fra nysgerrighed. Dermed kan tilstanden siges at fremme selvbestemt opmærksomhed. I tilstanden føles en indsats i forhold til læring - også en langvarig indsats - ubesværet. Og selv ved stor modstand og frustration vil en individuelt interesseret person udvise vedholdenhed og opsøgende adfærd. Ligesom i de tidligere faser af interesseudvikling er den individuelle interesse koblet til viden. Jo mere man ved og kan om et emne, jo mere interessant opleves det typisk at være (Hidi & Renninger, 2006).

Sammenfattende kan viden og positive følelser siges at være interdependente i udviklingen af interesse. Interesseudvikling involverer en interaktion mellem viden og positive følelser, og overgange mellem faser inkluderer både positive følelser og muligheder for at udvikle mere viden (Hidi & Renninger, 2006).

3.3 LIFEs teoretiske grundlag for måling af interesse og motivation

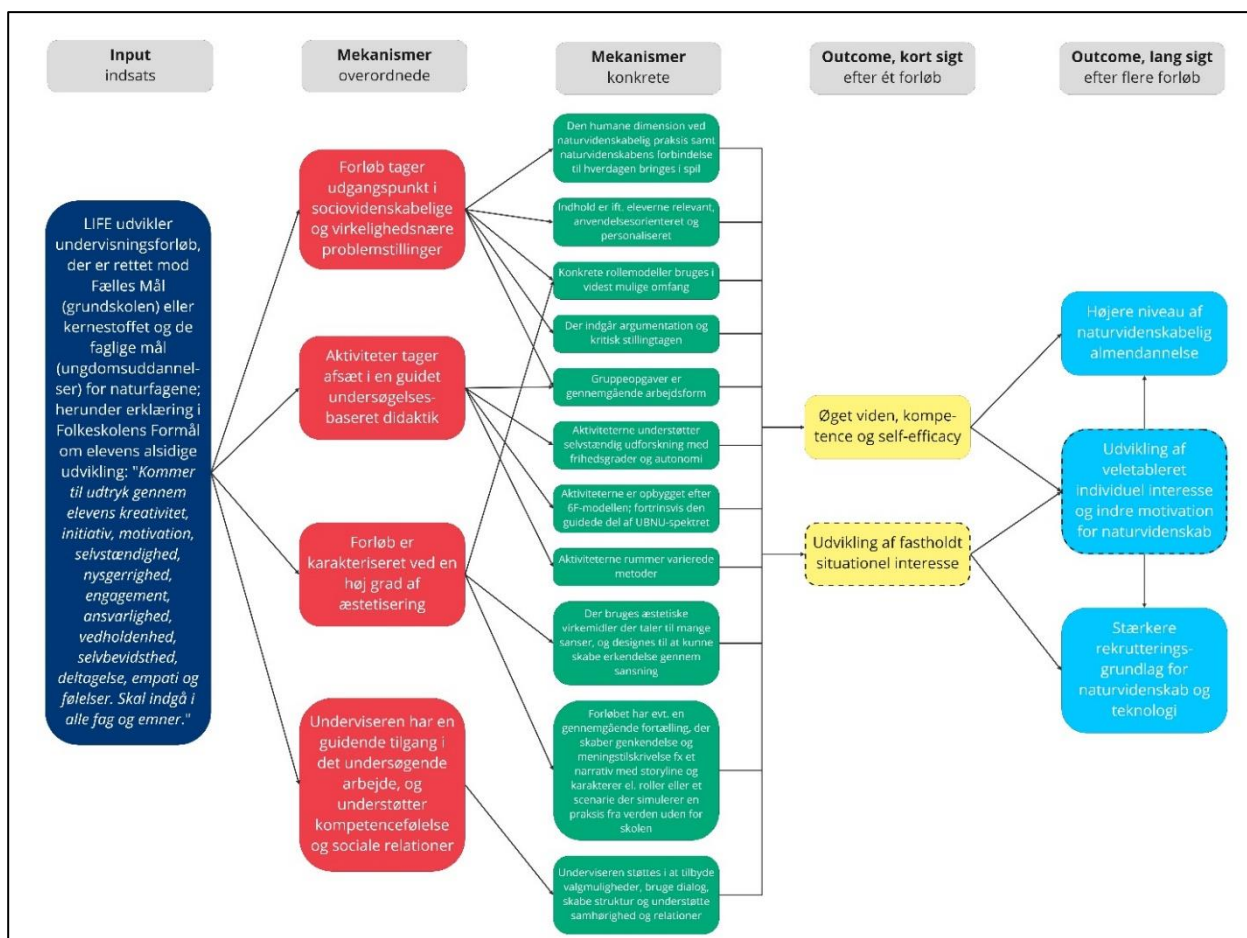
LIFEs didaktiske praksis er inspireret af Self-determination theory. Dette fremgår fx af LIFEs forandringsteori for undervisningsindsatser. Måling af missionsmålet om interesse og motivation tager følgelig afsæt i det teoretiske grundlag, som Self-determination theory hviler på, og som er beskrevet ovenfor. Derved skabes sammenhæng mellem det teorigrundlag, LIFEs undervisningsforløb bygger på, og den måde undervisningsforløbenes effekt på motivation og interesse måles på.

Endvidere tager målingen afsæt i den forståelse for interesseudvikling, som fire-fase-modellen hviler på, og som er beskrevet ovenfor. Det skyldes, at fire-fase-modellen er velegnet til at forstå, hvilke effekter vi kan forvente efter deltagelse i et enkelt LIFE Forløb som fx Turbovækst, og hvad vi kan forvente efter deltagelse i mange LIFE Forløb. Forståelsen i fire-fase-modellen om, at interesse udvikles i faser gående fra en flygtig situationel interesse til en veletableret individuel interesse, er meget brugbar i forhold til at opstille realistiske outcomemål om interesse og motivation på hhv. kort- og lang sigt (jf. de to stiplede gule/lyseblå bokse i Figur 3.3.1).

Teoretisk er der mange ligheder mellem veletableret individuel interesse i fire-fase-modellen og indre motivation i Self-determination theory. Lighedspunkterne er blandt andet, at høj værdsættelse, selvbestemt opmærksomhed og bevidste valg truffet ud fra nysgerrighed kendetegner en adfærd, der fremmes af individuel interesse. Og værdsættelse, selvbestemmelse og det at gøre noget for aktivitetens egen skyld ud fra en dyb interesse er netop kernen i indre motivation. Disse lighedspunkter er påpeget af en række forskere, fx Wigfield & Eccles (2000).

Lighedspunkterne har den konsekvens, at vi i udviklingen af måleinstrumenter for outcome på lang sigt kan lade os inspirere af måleinstrumenter baseret på Self-determination Theory såvel som fire-fase-modellen.

Figur 3.3.1. Forandringsteori for LIFEs undervisningsindsatser.



Note: De to stiplede bokse (hvh. gul og lyseblå) markerer hvilken effekt i forhold til interesse og motivation, der kan forventes på kort og lang sigt.

3.4 LIFE's måling af interesse og motivation

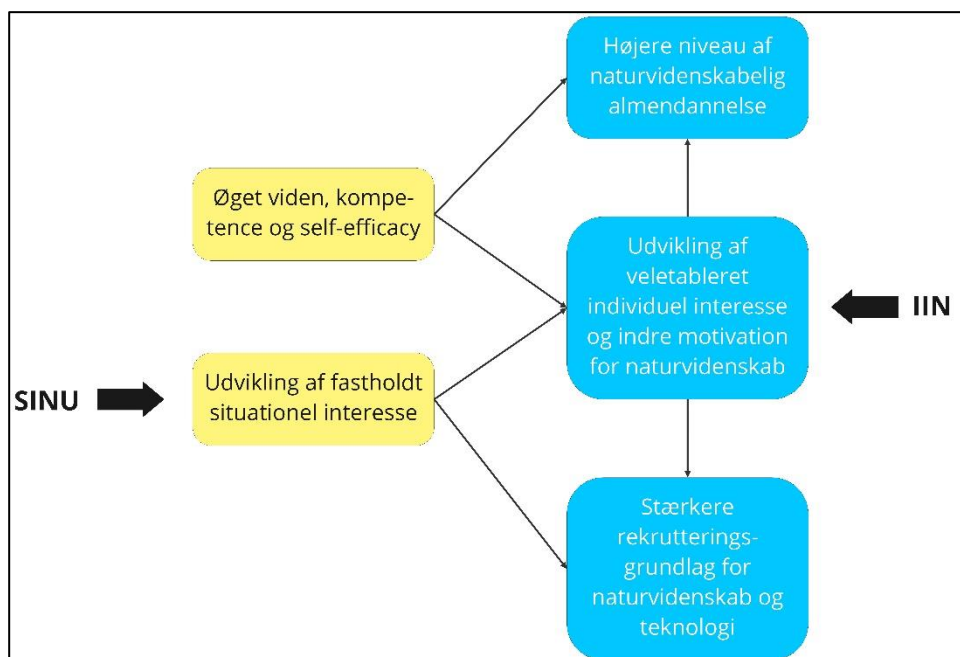
Ovenfor gav vi et overblik over de to teorier, der udgør det teoretiske grundlag for LIFE's måling af missionsmålet om interesse og motivation, nemlig Self-determination theory og fire-fase-modellen. I dette afsnit vil vi beskrive, hvordan vi bruger teorierne til at udvikle to måleinstrumenter til at måle elevernes interesse og motivation for naturfag. Først forklarer vi, hvordan måleinstrumenterne dels tager afsæt i LIFE's forandringsteori, dels i en række allerede udviklede survey-skalaer fra motivations- og naturfagsdidaktisk forskning. Derefter præsenterer vi de måleinstrumenter, LIFE bruger til at måle interesse og motivation.

Måleinstrumenternes sammenhæng med LIFE's forandringsteori

LIFE bruger to måleinstrumenter, der måler elevernes interesse og motivation for naturfag:

1. **Situationel interesse i naturfaglige undervisningsforløb (SINU).** Måleinstrumentet er udarbejdet med inspiration fra måleredskabet Situational Interest Measures (Linnenbrink-Garcia et al. 2010).
2. **Indre motivation for naturfag (IIN).** Måleinstrumentet er udarbejdet med afsæt i UiL-skalaen Indre motivation. Skalaen er valideret i en dansk skolekontekst (Makransky et al. 2020).

Figur 3.4.1. Markering af hvilke outcomemål i forandringsteorien (jf. Figur 3.3.1), de to måleinstrumenter måler.



Kilde: LIFE.

Første måleinstrument (SINU) måler den gule boks "udvikling af fastholdt situationel interesse". SINU er velegnet til at undersøge om de undervisningsaktiviteter, materialer og emner som eleverne møder i et LIFE-forløb som Turbovækst, kan vække en umiddelbar interesse for naturfag. Denne type interesse kan med afsæt i fire-fase modellen for interesseudvikling kaldes en situationel interesse. Det er en interesse, der opstår i elevernes møde med nogle specifikke læringsituationer som undervisningsforløbet Turbovækst. Det er væsentligt at måle, hvorvidt Turbovækst skaber en situationel interesse, da dette er forudsætningen for en eventuel udvikling af en vedvarende individuel interesse.

Det andet måleinstrument (IIN) måler den midterste lyseblå boks "Udvikling af veletableret individuel interesse og indre motivation for naturvidenskab". IIN er velegnet til at undersøge om Turbovækst er i stand til at øge elevernes vedvarende interesse for naturfag; med andre ord en interesse, der ikke er bundet op på specifikke læringsituationer, men som er udtryk for en generel og kontekstafhængig interesse for naturfagene. Med fire-fase modellen kan denne interesse kaldes en individuel interesse. Som tidligere påpeget har den individuelle interesse mange lighedspunkter med Self-determination theorys forståelse af indre motivation.

Det er LIFE's vurdering, at en relativt kortvarig indsats som Turbovækst har vanskeligt ved at påvirke elevernes individuelle interesse eller indre motivation for naturfag (målt med IIN). Det er derimod langt mere plausibelt, at et enkelt LIFE forløb potentielt kan påvirke elevernes situationelle interesse for naturfag (målt med SINU). På trods af dette vælger LIFE alligevel at afprøve, om et enkelt forløb kan afstedkomme en vedvarende interesse for naturfag hos eleverne. En medvirkende årsag til det valg er, at LIFE uanset resultatet vil opnå erfaringer med et måleinstrument, der kan bruges i senere effektmålinger, hvor effekten måles efter flere LIFE forløb.

De to måleredskaber er udarbejdet med inspiration fra den empiriske forskning i interesse og motivation. LIFE har gennemført et litteraturstudie, hvor vi har søgt efter forskningsartikler, der vha. survey har målt elevers motivation og interesse specifikt inden for naturfag. Nedenfor gennemgås først udarbejdelsen af måleredskabet SINU og dets

relation til andre anvendte forskningsbaserede måleinstrumenter. Dernæst gennemgås på samme måde udarbejdelsen af måleredskabet IIN.

Situational interesse i naturfaglige undervisningsforløb

Interesse målt i en undervisningskontekst måles på mange forskellige måder i forskningen. Et norsk studie giver et overblik over de mange forskellige måleinstrumenter, der er anvendt i forskningen (Dahl & Nierenberg 2021). I følge studiet har måleinstrumenterne ændret sig i takt med udviklingen i forståelsen af interesse. Før fire-fase-modellen blev formuleret, har måleinstrumenterne taget udgangspunkt i forskellige andre teoriretninger, fx expectancy-value teorien (Eccles et al., 1983), interesse som triggeret, situationelt og personligt (Schraw et al., 1995; Ainley et al., 2002), interesse som følelser og værdi (Krapp et al., 1988; Schiefele, 1999) samt self-concept-teorien (Marsh et al., 2005).

Der er fire surveybaserede måleinstrumenter, der er udviklet med afsæt i fire-fase-modellen (Dahl & Nierenberg 2021):

- Situational Interest Measures (SI) (Linnenbrink-Garcia et al. 2010)
- Individual Interest Questionnaire (IIQ) (Rotgans 2015)
- Situational Interest Scale (SIS) (Chen et al. 1999)³
- Motivationskalaer (Bathgate 2014)

LIFE vælger at udarbejde et måleinstrument, der primært er inspireret af måleinstrumentet Situational Interest Measures (SI). Som beskrevet ovenfor er måleinstrumentet teoretisk funderet i fire-fase-modellen og fokuserer eksplicit på at måle situational interesse. Måleinstrumentet er udviklet i to udgaver målrettet hhv. amerikanske bachelorstuderende og amerikanske elever i middle og high school, hvilket svarer til danske elever i 7.-9. klasse samt gymnasieelever. Sidstnævnte udgave er altså målrettet en gruppe af elever, der aldersmæssigt nogenlunde svarer til målgruppen for Turbovækst (9. klasseselever).

Situational Interest Measures er konstrueret til at måle fire-fase-modellens første to faser: igangsat situational interesse og fastholdt situational interesse. Igangsat situational interesse måles med fire items og fokuserer på elevernes affektive reaktion på læreren og undervisningsaktiviteterne. Fastholdt situational interesse måles med otte items, der måler elevernes følelsesmæssige reaktion både på det, de lærer, og på hvor meningsfuld aktiviteten opfattes. Forskellen på igangsat og fastholdt situational interesse er, at igangsat situational interesse fokuserer på den umiddelbare oplevelse af, om læreren og undervisningsaktiviteterne vækker interesse, mens fastholdt situational interesse fokuserer på om det, eleven lærer, opleves som interessant og meningsfuldt (Linnenbrink-Garcia et al. 2010).

Inspireret af SI er LIFE's måleinstrument udarbejdet, så der måles to underdimensioner: 1) igangsat situational interesse og 2) fastholdt situational interesse. Denne opdeling foretages, da sondringen empirisk har vist sig at have relevans (Linnenbrink-Garcia et al. 2010). Selvom LIFE er inspireret af items i SI, har der været behov for væsentlige tilpasninger. Vi har således udbygget måleinstrumentet med yderligere items og tilpasset det en dansk skolekontekst. Derudover har vi tilpasset måleinstrumentet, så der spørges til naturfagsundervisningen i dele af fremfor hele skoleåret.

Måleinstrumentet er konstrueret, så det er muligt at gennemføre som en før- eftermåling blandt elever, der *har* gennemført Turbovækst i efteråret 2023 (indsatsgruppen) og blandt elever, der *ikke har* gennemført Turbovækst i efteråret 2023 (kontrolgruppen). Herved er det muligt at undersøge Turbovæksts effekt på elevernes situationelle interesse for naturfag. Elever i kontrolgruppen bliver bedt om at vurdere den ordinære undervisning i naturfag.

³ Er udviklet før teorien om fire-fase-modellen, men er kraftigt inspireret af Hedi & Rennigers (2006) forståelse af situational interesse.

Herved måler måleinstrumentet om Turbovækst vækker en større eller mindre situationel interesse end den ordinære undervisning.

Tabel 3.4.1: Situationel interesse i naturfaglige undervisningsforløb

Igangsatt situationel interesse (4)

Hvor tit skete det at:

- Undervisningen var spændende
- Det vi *lavede* i undervisningen, gav mig lyst til at deltage
- De *materialer* vi brugte i undervisningen, gav mig lyst til at deltage
- De *emner* vi arbejdede med i undervisningen, gav mig lyst til at deltage

Fastholdt situationel interesse (4)

Hvor tit skete det at:

- Det, vi lærte om, var interessant
- Jeg tænkte på det, jeg havde lært efter skoletid
- Jeg oplevede, at det jeg lærte, kunne bruges uden for skolen
- Jeg kunne bruge det jeg lærte til bedre at forstå ting i min hverdag

Kilde: LIFE.

Indre motivation for naturfag

Som tidligere beskrevet, er der store sammenfald mellem det, Self-determination theory kalder *indre motivation* og det, fire-fase-modellen kalder *individuel interesse*.

Litteraturstudiet viser, at der i forskningen anvendes mange forskellige måleinstrumenter til at måle indre motivation og individuel interesse. Fire måleredskaber anvendes ofte i forskningen:

- Academic Self-Regulation Questionnaire (SRQ_A)
- Motivated Strategies for Learning Questionnaire (MSLQ)
- Science Motivation Questionnaire (SMQ)
- Individual Interest Questionnaire (IIQ)

Hvor de første tre måler indre motivation, måler det sidste individuel interesse. Academic Self-Regulation Questionnaire (SRQ_A) er udviklet af Ryan & Connell (1989) og måler de forskellige typer af motivation, der er knyttet til Self-Determination Theory, herunder indre motivation. Motivated Strategies for Learning Questionnaire (MSLQ) er udarbejdet af Pintrich et al. (1991) og måler en lang række dimensioner, der er vigtige for læring. Måleinstrumentet måler blandt andet *intrinsic* og *extrinsic goal orientation*, som i en lang række forskningsartikler er fortolket som indre og ydre motivation (se fx Nielsen, 2018). Det teoretiske ophav hviler på Banduras socialkognitive teori, men har overlap med Self-determination theorys motivationsformer. Det tredje anvendte måleinstrument er Science Motivation Questionnaire (SMQ) udviklet af Glynn & Koballa (2006), der måler fem dimensioner, herunder indre motivation samt to former for ydre motivation, navnlig *grade motivation* og *career motivation*. Det teoretiske ophav til måleinstrumentet er mere uklart, men det trækker både på Pintrichs videreudvikling af Banduras socialkognitive teori og på Self-determination theory (Glynn et al., 2009). Det fjerde og sidste måleinstrument, Individual Interest Questionnaire (IIQ), er udviklet af Rotgans (2015) og måler individuel interesse med afsæt i 4-fase-modellen for interesseudvikling.

På trods af de fire måleinstrumenters forskellige teoretiske afsæt er der klare fællestræk mellem dem. Indre motivation (fra Self-determination theory) og individuel interesse (fra fire-fase-modellen for interesseudvikling) bliver målt relativt ens på tværs af måleinstrumenterne, omend items er formuleret forskelligt. Indre motivation og individuel interesse måles konsekvent som noget, hvor eleven er drevet af nysgerrighed og interesse for naturfag, hvilket står i kontrast til mål for ydre motivation, hvor der måles om eleven er drevet af fx at få gode karakter. Nogle måleinstrumenter er udelukkende rettet mod undervisningen i naturfag i skolen og på uddannelsen (fx SRQ_A,

MSLQ og SMQ), mens andre måleinstrumenter også inddrager spørgsmål om interesse og motivation for naturvidenskab i fritiden (fx IIQ).

LIFE har valgt at tage udgangspunkt i de to forskningsbaseret mål for indre og ydre motivation udarbejdet af Nordic Metrics. Vi har ikke en forventning om at Turbovækst har en effekt for ydre motivation, men udelukkende har en effekt for den indre motivation. Det skyldes at LIFE ikke direkte i forløbene arbejder med at styrke ydre motivation. Ydre motivation er dog medtaget, da det er interessant at se, om LIFE forløb utilsigtet påvirker den ydre motivation.

De to mål for indre og ydre motivation er to ud af i alt 19 UiL-skalaer fra Nordic Metric, der måler elevernes udvikling i livsfærdigheder. De to mål er inspireret af MSLQ men tilpasset til en dansk grundskolekontekst og valideret blandt danske elever i 4.-9. klasse (Makransky et al. 2020⁴). De to mål er ikke målrettet specifikke fag. LIFE har derfor lavet nogle få tilpasninger, så items er målrettet Fysik/kemi. Fx er spørgsmålet i UiL-skalaen for indre motivation "Jeg kan godt lide udfordrende aktiviteter, så jeg kan lære nye ting", ændret til: "Jeg kan godt lide, at det, vi laver i fysik/kemi, udfordrer mig, så jeg kan lære nye ting". Bemærk at vi ikke anvender ordet "aktivitet", da flere elever i pilottesten af skemaet, forbandt aktiviteter med en fysisk aktivitet, som fx *ikke* kunne indebære at se en animation eller læse en fagtekst.

På baggrund af den statistiske validering (se kapitel 4) ekskluderede vi to spørgsmål fra skalaen der måler indre motivation og et spørgsmål fra skalaen der måler ydre motivation. Den statistiske validering viste at spørgsmålene ikke passede til skalaerne. Nedenfor ses de to måleinstrumenter.

Tabel 3.4.2: Motivation for naturfag

Indre motivation (3)

Er du enig eller uenig i følgende udsagn:

- Jeg kan godt lide, at det, vi laver i fysik/kemi, udfordrer mig, så jeg kan lære nye ting
 - Jeg kan godt lide, at det, vi laver i fysik/kemi, gør mig nysgerrig, også selvom det er svært
 - Jeg kan godt lide at lave noget i fysik/kemi, jeg lærer meget af
-

Ydre motivation (3)

Er du enig eller uenig i følgende udsagn:

- Det er vigtigt for mig at få en højere karakter end de andre i fysik/kemi
 - Jeg vil gerne klare mig godt i fysik/kemi, så familie og venner kan se, at jeg har succes
 - Det vigtigste for mig lige nu er at klare mig fagligt godt i fysik/kemi, så jeg får en god karakter
-

⁴ I Appendiks B i Makransky et al. (2020) ses spørgsmålene, der måler de to UiL-skalaer.

4. STATISTISK VALIDERING

I dette kapitel fremlægges resultaterne af skalavalideringen af de syv surveyskalaer præsenteret ovenfor. De syv surveyskalaer måler overordnet set to begreber: science self-efficacy (tre skalaer) og interesse og motivation for naturfag (fire skalaer). Skalavalideringen baserer sig på data fra 25 9.klasser, der i perioden april-marts 2023 har gennemført Tubovækst Lab i LIFE's laboratorier. Spørgeskemaet er besvaret af 310 elever med en svarprocent på 72. Skalavalideringen er gennemført med sparring fra Ann-Sophie Buchardt, cand.stat., ph.d., Biostatistisk Afdeling, Institut for Folkesundhedsvidenskab, Københavns Universitet. Ann-Sophie Buchardt har omfattende erfaring med skalavalidering og har givet sparring i forhold til analysestrategi samt gennemgået dokumentation af skalaerne.

Kapitlet indeholder fire afsnit. I første afsnit gives en kort introduktion til den valgte valideringsmetode. I afsnit to redegør vi for indsamlingen af de data, valideringen er foretaget på. I afsnit tre fremlægges de overordnede resultater af valideringsanalysen, hvorefter afsnit fire giver en detaljeret dokumentation for hver enkelt skala.

4.1 Rasch-modellen

For at undersøge validiteten af de syv skalaer så grundigt som muligt, har vi gennemført detaljerede item respons-analyser. Vi har benyttet Rasch-modellen, fordi en skala, som passer til Rasch-modellen, har optimale måleegenskaber og har en statistisk sufficient sumscore (Kreiner, 2009).

At sumscoren er statistisk sufficient betyder, at den indeholder al den nødvendige information for at kunne vurdere elevens niveau af det, der måles, og at vi derfor kan benytte sumscoren som skalascore i effektanalyserne. Det er kun niveauet af det, vi måler, fx indre motivation, der er bestemmende for, hvordan der svares på hvert enkelt spørgsmål i skalaen. Den samlede score på den pågældende skala vil derfor i sig selv indeholde tilstrækkelig information om elevens niveau i forhold til fx indre motivation.

I analyserne undersøger vi, om der er systematiske afvigelser i skalascoren, som er forårsaget af, at et eller flere spørgsmål ikke fungerer på samme måde for alle eleverne – det som kaldes differentiell item funktion (DIF). Hvis der er DIF, betyder det, at elever, der fx scorer lige højt på skalaen for indre motivation, ikke har opnået denne score, kun fordi de har en lige høj indre motivation, men også pga. andre karakteristika som fx køn eller fagligt niveau. Den skævhed i scoren, der opstår på grund af DIF, kan betyde, at efterfølgende statistiske analyser, hvor undergrupper af elever sammenlignes, også bliver biased. Vi har undersøgt DIF ud fra to baggrundsvariable: køn og standpunktskarakter.⁵

I analyserne undersøger vi også, om nogle spørgsmål i skalaen hænger stærkere sammen, end hvad der kan forklares af det, skalaen måler. Vi kan altså finde ud af, om det blot er niveauet af skalaen for fx indre motivation, der afgør, hvad der svares på et givent spørgsmål, eller om det også afhænger af, hvad der er svaret på et eller flere af de andre spørgsmål i skalaen. At items, der definerer skalaen, er betinget uafhængige givet den bagvedliggende faktor, der ikke er direkte observerbar i data, og dermed er latent, kaldes lokal uafhængighed og er et krav i alle almindelige målingsmodeller. Hvis der ikke tages højde for lokal afhængighed mellem spørgsmål, når reliabiliteten beregnes (fx ved beregning af Cronbach's alpha), vil reliabiliteten fremstå som højere, end den egentlig er.

I de tilfælde, hvor der er DIF og/eller lokal afhængighed, har vi benyttet den grafiske loglineære Rasch-model, hvor der kan tages højde for DIF og lokal afhængighed. I tilfælde af DIF, er det muligt efterfølgende at justere den samlede score for skævheder forårsaget af DIF og derved opnå en mere retvisende skalascore, der kan bruges på tværs af undergrupper (Kreiner, 2007). Dette har dog ikke været nødvendigt i nærværende skalavalidering, da der ikke er observeret DIF i de syv skalaer. I flere af skalaerne er der tilfælde af lokal afhængighed mellem spørgsmål.

⁵ De to baggrundsvariable har følgende kategorier: køn(dreng, pige, andet) og standpunktskarakter(12,10,7,4,2/0/-3).

Her har vi anvendt test-retest reliabilitetsmålet til beregning af reliabiliteten, som kan tage højde for dette og derved give et retvisende billede af skalaernes reliabilitet.

4.2 Datagrundlag

LIFE har i marts-april 2023 gennemført en dataindsamling af surveyinstrumentet til impactmåling af Turbovækst. Den praktiske gennemførelse beskrives nærmere herunder.

Dataindsamling

LIFE har gennemført en dataindsamling blandt samtlige elever, der i perioden 22. marts til 25. april 2023 har gennemført Turbovækst Lab. Spørgeskemaet besvares ved at scanne en QR-kode på telefon eller iPad eller ved at kopiere en webadresse på computer. I testperioden har 29 klasser gennemført Turbovækst Lab. Af de 29 klasser har 25 afgivet minimum én besvarelse. Dermed har fire klasser ikke afgivet nogen besvarelser. I de 29 klasser går der 576 elever. Det er det tal, læreren har oplyst i forbindelse med booking af forløbet. I de 25 klasser, der har minimum én besvarelse, går der 520 elever.

373 elever har besvaret undersøgelsen. Af dem har 362 besvaret hele spørgeskemaet, mens 11 kun har besvaret nogle spørgsmål. Holdes besvarelserne op imod det samlede antal klasser, der har gennemført labforløbet i perioden (dvs. 29), svarer det til en svarprocent på 65. Holdes det i stedet op imod de klasser, der har afgivet minimum én besvarelse, svarer det til en svarprocent på 72.

Udfordringer med svarvarighed

Eleverne har i gennemsnit brugt 6 minutter og 6 sekunder på at besvare spørgeskemaet. Dette tal opgøres som tiden i sekunder fra de scanner QR-koden og åbner spørgeskemaet, til de klikker på 'Afslut' i slutningen. I påskeferien 2023 lavede LIFE en justering i formatet for besvarelsen af spørgeskemaet. Inden den 11/4 2023 var der ikke sat nogen klar tidsramme for besvarelse af spørgeskemaet. Gennem løbende monitorering af svartider, kunne vi konstatere at mange elever brugte relativt kort tid på besvarelse af spørgeskemaet. For at tilskynde eleverne til at svare grundigere, indførte vi fra og med den 11/4 2023 en tydeligere tidsramme om besvarelse af spørgeskemaet. Herefter blev eleverne informeret om, at det tager 13 minutter at besvare spørgeskemaet, og at eleverne skal blive siddende på deres plads imens. Den gennemsnitlige svartid i perioden 22/3-10/4 2023 (259 respondenter) var 6 minutter, mens den i perioden 11/4-25/4 2023 (103 respondenter) er 6 minutter og 30 sekunder. Det tyder således på, at den ændrede instruktion har forhøjet svartiden med et halvt minut.

LIFE har frasortet elever der har besvaret spørgeskemaet på under 4 minutter. Der er i alt frasortet 52 besvarelser (14 %). Valideringsanalysen er derfor lavet på baggrund af 310 elevbesvarelser. Begrundelsen for at frasortere elever med svartider under 4 minutter er, at LIFE har testet, at det ikke er muligt både at læse og forholde sig til svarmuligheder på under 4 minutter. Derudover har LIFE undersøgt fordelingen af svartider. Fordelingen af svartider indikerer ligeledes udfordringer med lave svartider, fx er der en overraskende stor andel elever med svartider på mellem 2-3 minutter (set i forhold til resten af fordelingen).

De centrale fakta om dataindsamlingen er opsummeret i nedenstående Tabel.

Tabel 4.2.1: Dataindsamlingen

Resultat	Beskrivelse
29 klasser	Forventes at have gennemført labforløbet
25 klasser	Har afgivet mindst én besvarelse

576 elever	Går i de 29 klasser, der forventes at have gennemført labforløbet
520 elever	Går i en af de 25 klasser, der har afgivet minimum én besvarelse
373 elever	Har besvaret hele eller dele af spørgeskemaet
362 elever	Har besvaret hele spørgeskemaet
310 elever	Har besvaret spørgeskemaet med svartid på 4 minutter eller derover
65% af eleverne	Har besvaret spørgeskemaet (ud af alle klasserne)
72% af eleverne	Har besvaret spørgeskemaet (ud af klasserne med minimum én besvarelse)

4.3 Resultater af valideringsanalyser

Analyserne af grafiske loglineære Rasch-modeller blev foretaget ved hjælp af programmet DIGRAM (Kreiner & Nielsen, 2013). Resultaterne af valideringsanalyserne viser, at det er muligt at danne syv skalaer, der samlet set måler elevernes self-efficacy, interesse og motivation. Hver skala består af mellem 3 og 6 spørgsmål, og samlet set måles de syv skalaer ved 29 spørgsmål. I nedenstående tabel gives et overblik over, om der er lokal afhængighed mellem spørgsmål, om spørgsmål fungerer ens for undergrupper (DIF), reliabiliteten af skalaen samt skalaens korrelation med standpunktskarakter.

Tabel 4.3.1: Resultater af valideringsanalysen

Skala (antal spørgsmål)	Antal lokal afhængighed mellem spørgsmål	Antal items der har DIF	Reliabilitet	Korrelation med standpunktskarakter
Self-efficacy – Eksperimenter (5)	0	0	0,90	0,46
Self-efficacy – læringsmål kit (6)	0	0	0,83	0,46
Self-efficacy – læringsmål lab (4)	1	0	0,71	0,46
Igangsæt situationel interesse (4)	1	0	0,84	0,14
Fastholdt situationel interesse (4)	1	0	0,79	0,18
Indre motivation (3)	0	0	0,85	0,37
Ydre motivation (3)	1	0	0,79	0,22

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Skalaens korrelation med standpunktskarakter er beregnet ved gammakorrelationer.

Af tabellen ses det, at der er observeret lokal afhængighed mellem spørgsmål i 4 ud af 7 skalaer. Lokal afhængighed mellem spørgsmål har betydning for skalaens reliabilitet, men har ikke yderligere betydning for skalaens anvendelse. I afsnittet nedenfor kan man se hvilke spørgsmål, der er lokalt afhængige. I to tilfælde har vi undladt at inddrage et lokal afhængighedsled i den grafiske loglineære Rasch-model, da korrelationen var svagt negativ (gammaværdi < -0,1).

Tabellen viser også, at der ikke er fundet DIF i nogen af de syv skalaer, det betyder at skalaerne fungerer på samme måde for elever med forskelligt køn og fagligt niveau (målt som standpunktskarakter).

Reliabiliteten er for alle skaler over 0,7, som ofte fremhæves som en acceptabel grænseværdi (Kaplan & Saccuzzo 2017; Brennan 2006). Der er korrigeret for lokal afhængighed i udregningen af skalaernes reliabilitet.

Af tabellen ses at særligt de tre skalaer for self-efficacy korrelerer højt med elevernes standpunktskarakter givet i 9. klasse. Denne korrelation er i overensstemmelse med forskning i self-efficacy. Forskning, viser at elevens self-efficacy på tværs af fag har en væsentlig betydning for elevernes senere præstationer (se fx Marsh et al. 2018; Richardson et al. 2012; Honicke & Broadbent 2015). Zoomer man ind på det forskningsfelt, der mere specifikt undersøger self-efficacy inden for naturfag, er der ligeledes evidens for, at science self-efficacy har en betydning for præstationer (herunder karakter) i naturfag (Ferla et al. 2009; Britner & Pajares, 2006).

Af tabellen ses også, at indre motivation og i lidt mindre grad ydre motivation korrelerer med standpunktskarakter, mens der er en noget lavere korrelation mellem standpunktskarakter og de to skalaer for situationel interesse. Disse forskelle giver god mening, da særligt indre motivation for naturfag, er en mere stabil motivation for naturfag, som i højre grad må forventes at korrelere med standpunktskarakter, end den situationelle interesse, der kun måler interesse for en specifik læringssituation. Forskningen viser, at det særligt er indre motivation som har en effekt for elevernes karakter, og at ydre motivation i mindre grad har effekt for elevens karakter (Taylor et al. 2014). Der er langt mindre forskning i den situationelle interesses betydning for elevernes karakter.

4.4 Dokumentation af de enkelte skalaer

I dette afsnit fremstilles dokumentationen for de enkelte skalaer. Afsnittet indeholder grafiske fremstillinger af skalaerne, globale homogenitetstest samt test af DIF. Først fremstilles de tre skalaer der måler self-efficacy, dernæst fremstilles de fire skalaer der måler interesse og motivation.

Self-efficacy – eksperimenter

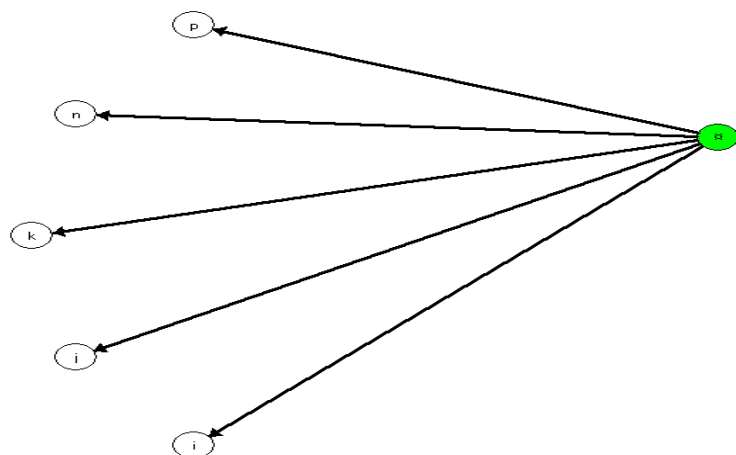
Skalaen måles med fem spørgsmål. Vurder, hvor sikker du er på, at du kan:

1. Lave en hypotese om, hvad du tror, resultatet af forsøget bliver (i)
2. Fortælle, hvilke variable (fx temperatur) du ændrer i forsøget, og hvilke variable du ikke ændrer i forsøget (j)
3. Gennemføre dataindsamling/målinger, du kan bruge til at vurdere resultatet af forsøget (k)
4. Forklare, hvad forsøget viser (n)
5. Forklare, hvordan forsøget gør dig klogere på det emne, du har om i fysik/kemi (p)

Svarkategorierne for de fem spørgsmål er: "Det kan jeg helt sikkert ikke" (1), "Det tror jeg ikke jeg kan" (2), "50 % sikker på jeg kan" (3), "Det tror jeg godt jeg kan" (4) og "Det kan jeg helt sikkert godt" (5).

Skalaen passer til en Rasch-model, hvor der hverken er DIF eller lokal afhængighed mellem spørgsmålene i skalaen. Reliabiliteten for skalaen er 0,90. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.1: Grafisk fremstilling af skalaen Self-efficacy - eksperimenter



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne i-p er de fem spørgsmål i skalaen.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.1: Den globale homogenitetstest samt globale test for DIF (n = 310)

Test	CLR	DF	p-værdi
Global homogenitetstest	19,4	19	0,431
Globale test for DIF:			
Køn	57,4	38	0,023
Standpunktskarakter	66,8	76	0,765

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$

Self-efficacy – læringsmål (kit)

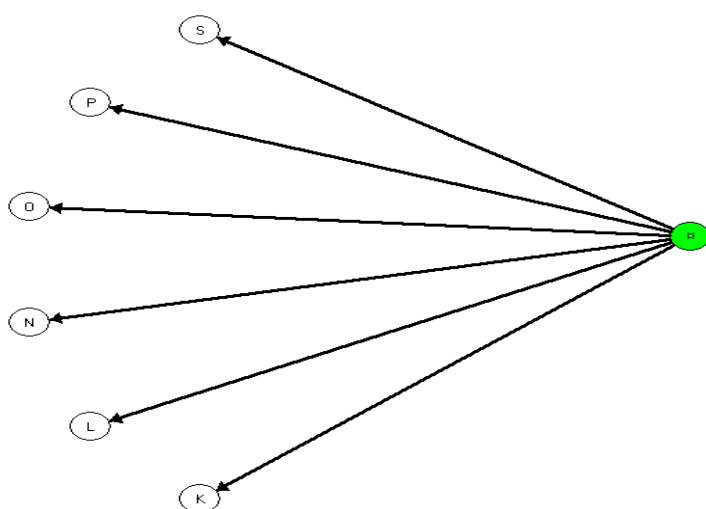
Skalaen måles med seks spørgsmål. Vurder, hvor sikker du er på, at du kan:

1. Fortælle, hvad et lands befolkningspyramide viser (K)
2. Fortælle, hvad en katalysator er (L)
3. Lave et forsøg, der viser om et metal er katalysator i en kemisk reaktion (N)
4. Lave et forsøg, der viser, om kunstgødning har betydning for planters vækst (O)
5. Give eksempler på både fordele og ulemper ved at bruge kunstgødning (P)
6. Give eksempler på, hvad et land kan gøre for at øge fødevarerproduktionen (S)

Svarkategorierne for de fem spørgsmål er: "Det kan jeg helt sikkert ikke" (1), "Det tror jeg ikke jeg kan" (2), "50 % sikker på jeg kan" (3), "Det tror jeg godt jeg kan" (4) og "Det kan jeg helt sikkert godt" (5).

Skalaen passer til en Rasch-model, hvor der hverken er DIF eller lokal afhængighed mellem spørgsmålene i skalaen. Reliabiliteten for skalaen er 0,83. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.2: Grafisk fremstilling af skalaen self-efficacy – læringsmål (kit)



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne K,L,N,O,P,S er de seks spørgsmål i skalaen.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.2: Den globale homogenitetstest samt globale test for DIF (n = 310)

Test	CLR	DF	p-værdi
Global homogenitetstest	17,7	23	0,775
Globale test for DIF:			
Køn	44,9	46	0,520
Standpunktskarakter	101,1	92	0,241

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$

Self-efficacy – læringsmål (lab)

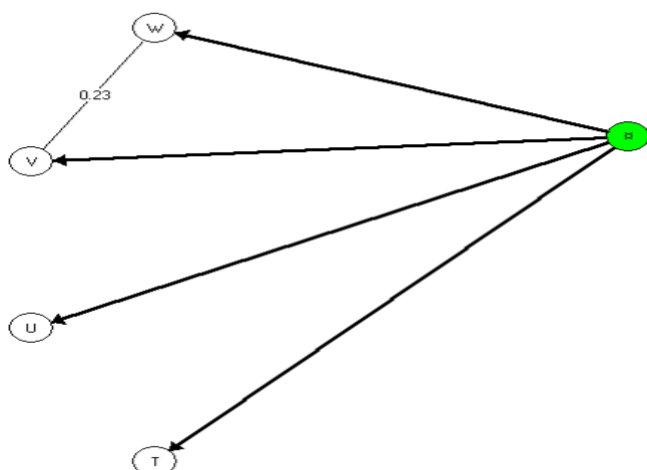
Skalaen måles med fire spørgsmål. Vurder, hvor sikker du er på, at du kan:

1. Give et eksempel på, hvad ammoniak kan bruges til (T)
2. Skrive ammoniak som en kemisk formel (U)
3. Nævne mindst to variable, der er vigtige, når et anlæg producerer ammoniak (V)
4. Lave et forsøg, der viser, om temperatur er vigtig for produktion af ammoniak (W)

Svarkategorierne til de fire spørgsmål er: "Det kan jeg helt sikkert ikke" (1), "Det tror jeg ikke, jeg kan" (2), "50% sikker på, jeg kan" (3), "Det tror jeg godt, jeg kan" (4) og "Det kan jeg helt sikkert godt" (5).

Skalaen passer til en grafisk loglineær Rasch-model, hvor der er lokal afhængighed mellem spørgsmålene V og W i skalaen. Der er ingen DIF mellem spørgsmålene i skalaen og de to baggrundsspørgsmål. Reliabiliteten for skalaen er 0,71. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.3: Grafisk fremstilling af skalaen self-efficacy - læringsmål (lab)



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne T-W er de fire spørgsmål i skalaen. Tallet mellem spørgsmålene beskriver korrelationen vha. standardiseret gammakoefficient.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.3: Den globale homogenitetstest samt globale test for DIF (n = 310)

Test	CLR	DF	p-værdi
Global homogenitetstest	28,5	28	0,436
Globale test for DIF:			
Køn	68,8	56	0,118
Standpunktskarakter	107,5	112	0,602

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$

Igangsæt situationel interesse

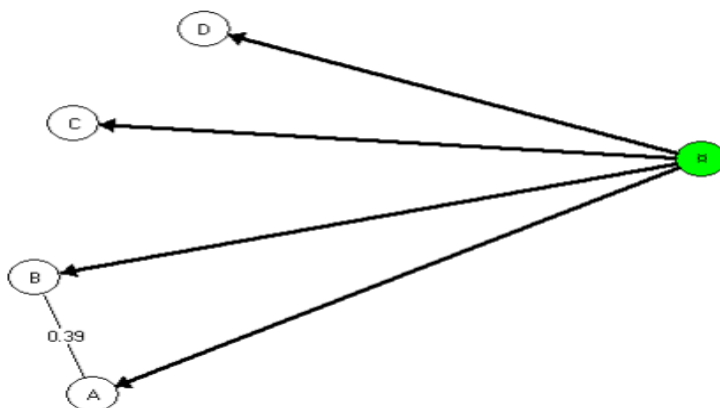
Skalaen måles med fire spørgsmål. Hvor tit skete det, at:

1. Undervisningen var spændende (A)
2. Det, vi lavede i undervisningen, gav mig lyst til at deltage (B)
3. De materialer, vi brugte i undervisningen, gav mig lyst til at deltage (C)
4. De emner, vi arbejdede med i undervisningen, gav mig lyst til at deltage (D)

Svarkategorierne til de fire spørgsmål er: "Aldrig" (1), "Sjældent" (2), "Nogle gange" (3), "Ofte" (4) og "Hver lektion/time" (5).

Skalaen passer til en grafisk loglineær Rasch-model, hvor der er lokal afhængighed mellem spørgsmålene A og B i skalaen. Der er ingen DIF mellem spørgsmålene i skalaen og de to baggrundsspørgsmål. Reliabiliteten for skalaen er 0,84. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.4: Grafisk fremstilling af skalaen Igangsæt situationel interesse



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne A-D er de fire spørgsmål i skalaen. Tallet mellem spørgsmålene beskriver korrelationen vha. standardiseret gammakoefficient.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.4: Den globale homogenitetstest samt globale test for DIF ($n = 310$)

Test	CLR	DF	p-værdi
------	-----	----	---------

Global homogenitetstest	18,6	25	0,816
Globale test for DIF:			
Køn	51,6	50	0,409
Standpunktskarakter	107,6	100	0,283

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$.

Fastholdt situationel interesse

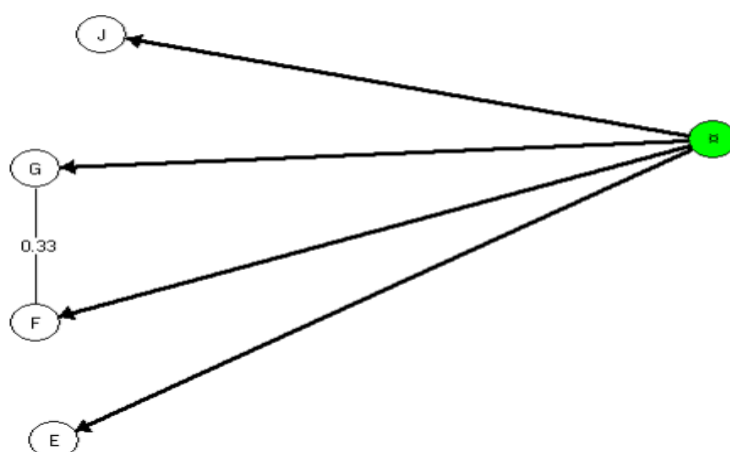
Skalaen måles med fire spørgsmål. Hvor tit skete det, at:

1. Det, vi lærte om, var interessant (E)
2. Jeg tænkte på det, jeg havde lært, efter skoletid (F)
3. Jeg oplevede, at det jeg lærte, kunne bruges uden for skolen (G)
4. Jeg kunne bruge det, jeg lærte, til bedre at forstå ting i min hverdag (J)

Svarkategorierne til de fire spørgsmål er: "Aldrig" (1), "Sjældent" (2), "Nogle gange" (3), "Ofte" (4) og "Hver lektion/time" (5).

Skalaen passer til en grafisk loglineær Rasch-model, hvor der er lokal afhængighed mellem spørgsmålene F og G i skalaen. Der er ingen DIF mellem spørgsmålene i skalaen og de to baggrundsspørgsmål. Reliabiliteten for skalaen er 0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.5: Grafisk fremstilling af skalaen Fastholdt situationel interesse



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne E, F, G, J er de fire spørgsmål i skalaen. Tallet mellem spørgsmålene beskriver korrelationen vha. standardiseret gammakoefficient.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.5: Den globale homogenitetstest samt globale test for DIF (n = 310)

Test	CLR	DF	p-værdi
Global homogenitetstest	12,0	26	0,991
Globale test for DIF:			
Køn	64,5	52	0,114
Standpunktskarakter	115,2	104	0,213

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$.

Indre motivation

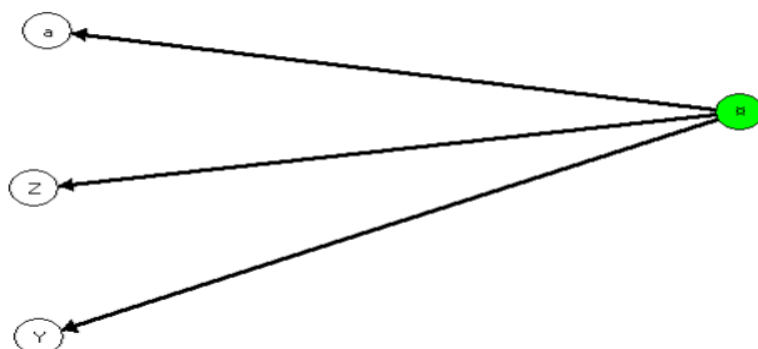
Skalaen måles med tre spørgsmål. Er du enig eller uenig i følgende udsagn:

1. Jeg kan godt lide, at det, vi laver i fysik/kemi, udfordrer mig, så jeg kan lære nye ting (Y)
2. Jeg kan godt lide, at det, vi laver i fysik/kemi, gør mig nysgerrig, også selvom det er svært (Z)
3. Jeg kan godt lide at lave noget i fysik/kemi, jeg lærer meget af (a)

Svarkategorierne til de tre spørgsmål er: "Helt enig" (1), "Nærmest enig" (2), "Hverken/eller" (3), "Nærmest uenig" (4) og "Helt uenig" (5).

Skalaen passer til en Rasch-model, hvor der hverken er DIF eller lokal afhængighed mellem spørgsmålene i skalaen. Reliabiliteten for skalaen er 0,85. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.6: Grafisk fremstilling af skalaen Indre motivation



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne Y, Z, a er de tre spørgsmål i skalaen.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.6: Den globale homogenitetstest samt globale test for DIF (n = 310)

Test	CLR	DF	p-værdi
Global homogenitetstest	8,1	11	0,709
Globale test for DIF:			
Køn	36,5	22	0,027
Standpunktskarakter	44,1	44	0,467

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$.

Ydre motivation

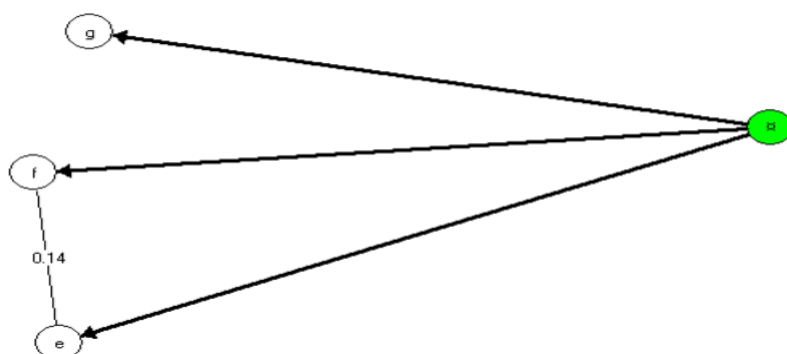
Skalaen måles med tre spørgsmål. Er du enig eller uenig i følgende udsagn:

1. Det er vigtigt for mig at få en højere karakter end de andre i fysik/kemi (e)
2. Jeg vil gerne klare mig godt i fysik/kemi, så familie og venner kan se, at jeg har succes (f)
3. Det vigtigste for mig lige nu er at klare mig fagligt godt i fysik/kemi, så jeg får en god karakter (g)

Svarkategorierne til de tre spørgsmål er: "Helt enig" (1), "Nærmest enig" (2), "Hverken/eller" (3), "Nærmest uenig" (4) og "Helt uenig" (5).

Skalaen passer til en grafisk loglineær Rasch-model, hvor der er lokal afhængighed mellem spørgsmålene e og f i skalaen. Der er ingen DIF mellem spørgsmålene i skalaen og de to baggrundsspørgsmål. Reliabiliteten for skalaen er 0,79. I nedenstående figur ses en grafisk fremstilling af skalaen.

Figur 4.4.7: Grafisk fremstilling af skalaen Ydre motivation



Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: Den grønne cirkel er skalaen og cirklerne e-g er de tre spørgsmål i skalaen. Tallet mellem spørgsmålene beskriver korrelationen vha. standardiseret gammakoefficient.

Nedenstående tabel viser, at spørgsmålene på skalaen fungerer på samme måde for de elever, som scorer højt, og de, som scorer lavt på skalaen (global homogenitetstest). Tabellen viser også, at der ikke er DIF mellem spørgsmålene i skalaen og baggrundsvariablene køn og standpunktskarakter (Globale test for DIF).

Tabel 4.4.7: Den globale homogenitetstest samt globale test for DIF ($n = 310$)

Test	CLR	DF	p-værdi
Global homogenitetstest	20,7	26	0,758
Globale test for DIF:			
Køn	23,7	26	0,592
Standpunktskarakter	97,1	104	0,672

Kilde: Rasch-analyse af data fra pilotdataindsamling blandt 9. klasseelever, marts/april 2023.

Note: CLR er Conditional Likelihood Ratio-test, der tester, om spørgsmålsparametrene i modellen er ens i nærmere definerede grupper. Den globale homogenitetstest sammenligner spørgsmålsparametre for to lige store grupper af elever, hvor den ene gruppe scorer højt, og den anden gruppe scorer lavt på skalaen. De globale test for DIF sammenligner på samme måde, om spørgsmålsparametrene fungerer ens i de forskellige undergrupper defineret ved baggrundsvariablene. Signifikansniveauet korrigeres for, at der gennemføres flere test: *5 % signifikansniveau = $p < 0,0167$; **1 % signifikansniveau = $p < 0,0033$.

Note: Det globale test for DIF på køn konvergerede ikke. Det kan skyldes at der kun er 14 observationer i kategorien "andet køn". Tallene der ses i tabellen specifikt for test af DIF på køn, er dannet på et datagrundlag uden de 14 observationer. Vi har desuden testet om, der er DIF mellem køn og de tre spørgsmål i skalaen. Resultatet af analysen viser, at der ikke er DIF (p -værdier $\geq 0,219$).

5. SURVEYSKALAER

Self-efficacy – Eksperimenter (5)

Forestil dig, at du skal lave et forsøg i Fysik/kemi.

Vurder, hvor sikker du er på, at du kan:

- Lave en hypotese om, hvad du tror, resultatet af forsøget bliver (i)
- Fortælle, hvilke variable (fx temperatur) du ændrer i forsøget, og hvilke variable du ikke ændrer i forsøget (j)
- Gennemføre dataindsamling/målinger, du kan bruge til at vurdere resultatet af forsøget (k)
- Forklare, hvad forsøget viser (n)
- Forklare, hvordan forsøget gør dig klogere på det emne, du har om i fysik/kemi (p)

Self-efficacy – Læringsmål kit (6)

Vurder, hvor sikker du er på, at du kan:

- Fortælle, hvad et lands befolkningspyramide viser (K)
- Fortælle, hvad en katalysator er (L)
- Lave et forsøg, der viser om et metal er katalysator i en kemisk reaktion (N)
- Lave et forsøg, der viser, om kunstgødning har betydning for planters vækst (O)
- Give eksempler på både fordele og ulemper ved at bruge kunstgødning (P)
- Give eksempler på, hvad et land kan gøre for at øge fødevareproduktionen (S)

Self-efficacy – Læringsmål lab (4)

Vurder, hvor sikker du er på, at du kan:

- Give et eksempel på, hvad ammoniak kan bruges til (T)
- Skrive ammoniak som en kemisk formel (U)
- Nævne mindst to variable, der er vigtige, når et anlæg producerer ammoniak (V)
- Lave et forsøg, der viser, om temperatur er vigtig for produktion af ammoniak (W)

Igangsæt situationel interesse (4)

Hvor tit skete det, at:

- Undervisningen var spændende (A)
- Det, vi lavede i undervisningen, gav mig lyst til at deltage (B)
- De materialer, vi brugte i undervisningen, gav mig lyst til at deltage (C)
- De emner, vi arbejdede med i undervisningen, gav mig lyst til at deltage (D)

Fastholdt situationel interesse (4)

Hvor tit skete det, at:

- Det, vi lærte om, var interessant (E)
- Jeg tænkte på det, jeg havde lært, efter skoletid (F)
- Jeg oplevede, at det jeg lærte, kunne bruges uden for skolen (G)
- Jeg kunne bruge det, jeg lærte, til bedre at forstå ting i min hverdag (J)

Indre motivation (3)

Er du enig eller uenig i følgende udsagn:

- Jeg kan godt lide, at det, vi laver i fysik/kemi, udfordrer mig, så jeg kan lære nye ting (Y)
- Jeg kan godt lide, at det, vi laver i fysik/kemi, gør mig nysgerrig, også selvom det er svært (Z)
- Jeg kan godt lide at lave noget i fysik/kemi, jeg lærer meget af (a)

Ydre motivation (3)

Er du enig eller uenig i følgende udsagn:

- Det er vigtigt for mig at få en højere karakter end de andre i fysik/kemi (e)
- Jeg vil gerne klare mig godt i fysik/kemi, så familie og venner kan se, at jeg har succes (f)

- Elmose, S. (2015). Naturfaglig Kompetence: – baggrund for begrebet, dets styrker og begrænsninger i naturfagsundervisning. <https://astra.dk/blog/ntsadmin/kompetencer>
- Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19(4), 499–505. <http://dx.doi.org/10.1016/j.lindif.2009.05.004>.
- Fidan, M., & Tuncel, M. (2021). Developing a self-efficacy scale toward physics subjects for lower-secondary school students. *Journal of Baltic Science Education*, 20(1), 38–49. <https://doi.org/10.33225/jbse/15.14.311>
- Glynn, S. M., & Koballa, T. R., Jr. (2006). Motivation to learn college science. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.
- Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, 46, 127-146.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour economics*, 19(4), 451-464.
- Hidi, S. & Renninger, K.A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 111-127.
- Honick, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, 17, 63–84. <https://doi.org/10.1016/j.edurev.2015.11.002>
- Hu, X., Jiang, Y. & Bi, H. Measuring science self-efficacy with a focus on the perceived competence dimension: using mixed methods to develop an instrument and explore changes through cross-sectional and longitudinal analyses in high school. *IJ STEM Ed* 9, 47 (2022). <https://doi.org/10.1186/s40594-022-00363-x>
- Keilow, M., Brændegaard, N. W., Jæger, M. M., Pedersen, H. S. (2023). Udvikling af et redskab til måling af science-kapital – Teknisk rapport. København. Det Nationale Forsknings- og Analysecenter for Velfærd (VIVE)
- Krapp, A., Schiefele, U., and Winteler, A. (1988). Studieninteresse und fachbezogene Wissensstruktur. *Psychol. Erziehung Unterricht*. 35, 106–118
- Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12, 383-409.
- Kreiner, S., & Nielsen, T. (2013). *Item analysis in DIGRAM 3.04: Part I: Guided tours*. Department of Biostatistics, University of Copenhagen.
- Kreiner, S. (2009). Om udvikling og afprøvning af pædagogiske test. In C. Bendixen, & S. Kreiner (Eds.), *Test i folkeskolen* (pp. 43-82). Hans Reitzels Forlag.
- Lin, T.-J., & Tsai, C.-C. (2013). A multi-dimensional instrument for evaluating Taiwan high school students' learning self-efficacy in relation to their approaches to learning science. *International Journal of Science and Mathematics Education*, 11(6), 1275–1301. <https://doi.org/10.1007/s10763-012-9376-6>
- Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., et al. (2010). Measuring Situational Interest in Academic Domains. *Educ. Psychol. Meas.* 70 (4), 647–671. doi:10.1177/0013164409355699
- Makransky, G., Wandall, J., Madsen, S. R., Hood, M., & Creed, P. (2020). Development and Validation of the UiL-Scales for Measurement of Development in Life Skills—A Test Battery of Non-Cognitive Skills for Danish School

Children. *Scandinavian Journal of Educational Research*, 64(4), 612-627.

Doi:<https://doi.org/10.1080/00313831.2019.1595716>

Marsh H.W, Martin A.J, Yeung A.S, Craven R.G. (2018). Competence self-perceptions Elliott, C.S. Dweck, D.S. Yeager (Eds.), *Handbook of Competence and Motivation* (2nd ed.), Guilford Press (2018), pp. 85-115

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child. Dev.* 76 (2), 397–416. doi:10.1111/j.1467-8624.2005.00853.x

Nielsen, T. (2018). The intrinsic and extrinsic motivation subscales of the motivated strategies for learning questionnaire: a Rasch-based construct validity study. *Cogent Education*, 5, 1504485.

Petersen, M.R. (2017). Elevers interesse og motivation for STEM. In J.A. Nielsen (ed.), *Litteraturstudium til arbejdet med en national naturvidenskabsstrategi* (pp. 73-82). København: Institut for Naturfagenes Didaktik.

Pintrich, P.R., Smith, D.A.F., Garcia, T. & McKeachie, W.J. (1991). A manual for the use of the motivated strategies for learning questionnaire (MSLQ). Technical Report No. 91-8-004. The Regents of The University of Michigan.

Potvin, Patrice & Abdelkrim Hasni (2014) Interest, motivation and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research, *Studies in Science Education*, 50:1, 85-129, DOI: 10.1080/03057267.2014.881626

Reinhold, Sarah, Doris Holzberger & Tina Seidel (2018) Encouraging a career in science: a research review of secondary schools' effects on students' STEM orientation, *Studies in Science Education*, 54:1, 69-103, DOI: 10.1080/03057267.2018.1442900

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353. <http://dx.doi.org/10.1037/a0026838>

Riga, F., Winterbottom, M., Harris, E. & Newby, L. (2017). Inquiry-based science education. I K.S. Taber & B. Akpan (red.), *Science Education: an international course companion* (s. 247-261). Sense Publishers.

Rotgans, J. I. (2015). Validation Study of a General Subject-Matter Interest Measure: The Individual Interest Questionnaire (IIQ). *Health Professions Education*. 1 (1), 67–75. doi:10.1016/j.hpe.2015.11.009

Rüschenpöhler, Lilith & Silvija Markic (2019) Self-concept research in science and technology education – theoretical foundation, measurement instruments, and main findings, *Studies in Science Education*, 55:1, 37-68, DOI: 10.1080/03057267.2019.1645533

Ryan, R.M., & Connell, J.P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57, 749-761.

Ryan, R.M. & Deci, E.L. (2017). *Self-Determination Theory*. New York: The Guilford Press.

Ryan, R.M. & Deci, E.L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860.

- Sadi, Ozlem & Çakıroğlu, Jale. (2014). Relations of Cognitive and Motivational Variables with Students' Human Circulatory System Achievement in Traditional and Learning Cycle Classrooms. *Educational Sciences: Theory and Practice*. 14. 1197-2012. 10.12738/estp.2014.5.2037.
- Schiefele, U. (1999). Interest and Learning from Text. *Scientific Stud. Reading*. 3 (3), 257–279. doi:10.1207/s1532799xssr0303_4
- Schraw, G., Bruning, R., and Svoboda, C. (1995). Sources of Situational Interest. *J. Reading Behav.* 27 (1), 1–17. doi:10.1080/10862969509547866
- Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield, J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 15-31). San Diego, CA, US: Academic Press. doi:10.1016/B978-012750053-9/50003-6
- Summers, J. J., & Falco, L. D. (2022). Evaluating Construct Validity of the Middle School Self-Efficacy Scale With High School Adolescents. *Journal of Career Development*, 49(4), 735–752. <https://doi.org/10.1177/0894845318773383>
- Tatar, N., Yıldız, E., Akpınar, E., & Ergin, Ö. (2009). A study on developing a self-efficacy scale towards science and technology. *Eğitim Araştırmaları Eurasian Journal of Educational Research*, 36, 263–280.
- Tezer, M., & Aşıksoy, G. Y. (2015). Engineering students' self-efficacy related to physics learning. *Journal of Baltic Science Education*, 14(3), 311–326. <https://doi.org/10.33225/jbse/15.14.311>
- Tsai, M.-J., Wang, C.-Y., Wu, A.-H., & Hsiao, C.-Y. (2021). The Development and Validation of the Robotics Learning Self-Efficacy Scale (RLSES). *Journal of Educational Computing Research*, 59(6), 1056–1074. <https://doi.org/10.1177/0735633121992594>
- Tuan, H., Chin, C., & Shieh, S. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education*, 27, 639–654. <https://doi.org/10.1080/0950069042000323737>
- Usher, Ellen L. & Frank Pajares (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, Volume 34, Issue 1, 89-101.
- Uzuntiryaki, E., & Aydın, Y. Ç. (2009). Development and validation of chemistry self-efcacy scale for college students. *Research in Science Education*, 39(4), 539–551. <https://doi.org/10.1007/s11165-008-9093-x>
- Wigfield, A. & Eccles, J.S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81.
- Willis, G. B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications.

Bilag A1. Proces for udvikling af skalaer

Til at udvikle de tre skalaer for self-efficacy og de fire skalaer for interesse og motivation fulgte vi de syv trin anbefalet af Association for Medical Education in Europe (AMEE) (Artino et al., 2014).

Figur 2.5.1: Syv trin i udvikling af survey-skalaer

Step	Purpose
1. Conduct a literature review	To ensure that the construct definition aligns with relevant prior research and theory and to identify existing survey scales or items that might be used or adapted
2. Conduct interviews and/or focus groups	To learn how the population of interest conceptualizes and describes the construct of interest
3. Synthesize the literature review and interviews/focus groups	To ensure that the conceptualization of the construct makes theoretical sense to scholars in the field and uses language that the population of interest understands
4. Develop items	To ensure items are clear, understandable and written in accordance with current best practices in survey design
5. Conduct expert validation	To assess how clear and relevant the items are with respect to the construct of interest
6. Conduct cognitive interviews	To ensure that respondents interpret items in the manner that survey designer intends
7. Conduct pilot testing	To check for adequate item variance, reliability and convergent/discriminant validity with respect to other measures

Kilde: Artino et al. (2014).

I det følgende beskriver vi de enkelte trin i udviklingen af skalaerne.

Litteraturstudie (trin 1)

Med henblik på at udarbejde ovenstående surveyskalaer er der i LIFE foretaget en systematisk litteratursøgning for at sikre, at surveyskalaerne bygger på den nyeste og mest relevante forskning.

Litteraturstudiet har et dobbelt formål: 1) At få indsigt i det teoretiske grundlag for self-efficacy, interesse og motivation 2) At kortlægge, hvordan begreberne tidligere i forskningen er målt med surveyinstrumenter. Vi har opstillet en række søgekriterier inspireret af Rüschenpöhler og Markic (2019). Vi har valgt at gennemføre litteratursøgningen i databasen Education Resources Information Center (ERIC). ERIC er en fremtrædende og bredt anerkendt database for uddannelsesforskning, der indeholder alle de væsentlige tidsskrifter inden for uddannelse generelt og STEM-uddannelse specifikt (se f.eks. Rüschenpöhler & Markic, 2019; Potvin & Hasni, 2014; Reinhold et al., 2018). Vi har kun inkluderet fagfællebedømte artikler i en tiårsperiode (fra 2012 og frem). Vi har også inkluderet enkelte særligt citerede artikler udgivet før 2012.

Vi har begrænset søgningen til artikler, der fokuserer på uddannelse inden for naturvidenskab/STEM-emner. Derudover begrænsede vi resultaterne til artikler, der indeholder mindst et af ordene 'instrument', 'skala' og 'måling'. Dermed fanger vi den del af litteraturen, der er relevant for vores specifikke formål. Søgningen er yderligere begrænset til at omhandle grundskole og gymnasial uddannelse.

Ekspertvalidering og kognitive interviews (trin 2-6)

Inden udvikling af de konkrete items til de syv skalaer har vi gennemført interviews med tre 9. klasseselever og to forskere. Elevinterviewene gav viden om 9. klasseselevers sprogbrug og deres måde at italesætte mestrings- og interesse/motivationsoplevelser på. Forud for udvikling af items afholdtes møder med professor emeritus, Jens Dolin, fra Københavns universitet fra Institut for Naturfagernes Didaktik samt professor, Guido Makransky, fra Københavns universitet afdelingen for psykologi. Jens Dolin har indgående kendskab til naturfagsdidaktik, mens Guido Makransky har indgående kendskab til kvantitativ måling af self-efficacy. Sammen med resultaterne fra litteraturgennemgangen udviklede vi det første udkast til konkrete items til skalaerne. Efterfølgende gav eksperterne feedback på de enkelte items for at vurdere, hvor relevante items er i forhold til selv-efficacy begrebet. På samme vis er Morten Rask

Petersen, docent i anvendt matematik- og naturfagsdidaktik fra University College Lillebælt og Niels Bonnerup Dohn, lektor ved Aarhus universitet, inddraget i udviklingen af interesse- og motivationsskalaerne.

Efter eksperternes gennemgang af items gennemførte vi kognitive interviews med ni 9. klasseselever med inspiration fra Willis (2005). Interviewene blev gennemført som tre gruppeinterviews med tre elever ad gangen. I interviewene blev eleverne bedt om at læse et spørgsmål ad gangen og fortælle, hvor de vil placere sig på svarskalaen samt begrunde deres besvarelse. På denne måde er det muligt at teste om eleverne forstår spørgsmål og svarkategorier som intenderet og at identificere sprogbrug, der ikke er klart for målgruppen. På baggrund af de kognitive interviews blev der foretaget en række justeringer af items.

Pilottest og statistisk validering (trin 7)

I marts til april 2023 blev der gennemført en pilottest af spørgeskemaet. Spørgeskemaet blev besvaret af 310 9. klasser elever med en svarprocent på 65%. For at undersøge validiteten af skalaerne gennemførte vi detaljerede item-response-analyser ved Rasch-modellen. Fremgangsmåden og resultaterne af den statistiske validering fremgår af dette notats kapitel 4.