

De nationale tests samvariation med karakterer

Delrapport 2: Evalueringen af de nationale test



Peter Rohde Skov og Lasse Hønge Flarup

*De nationale tests samvariation med karakterer – Delrapport 2:
Evalueringen af de nationale test*

© VIVE og forfatterne, 2020

e-ISBN: 978-87-7119-741-9

Modelfoto: Ole Bo Jensen/VIVE

Projekt: 301403

VIVE – Viden til Velfærd

Det Nationale Forsknings- og Analysecenter for Velfærd

Herluf Trolles Gade 11, 1052 København K

www.vive.dk

VIVEs publikationer kan frit citeres med tydelig kildeangivelse.

Indhold

Sammenfatning	6
Resultater	6
1 Indledning	9
1.1 Formål	9
1.2 Læsevejledning	9
2 Design og metode	10
2.1 Rasch-modellen	10
2.2 Metode	13
2.3 Tidligere forskning	15
3 Dansk (læsning)	17
3.1 Dansk (læsning) i 2. klasse	17
3.2 Dansk (læsning) i 4. klasse	18
3.3 Dansk (læsning) i 6. klasse	18
3.4 Dansk (læsning) i 8. klasse	19
4 Matematik	22
4.1 Matematik i 3. klasse	22
4.2 Matematik i 6. klasse	23
5 Biologi i 8. klasse	25
6 Fysik/kemi i 8. klasse	26
7 Geografi 8. i klasse	27
8 Engelsk 7. i klasse	28
Litteratur	29
Bilag 1 Tabeller uden opdeling	31

Forord

Folketinget vedtog i marts 2006 indførelsen af de nationale test. Den første obligatoriske testrunde blev gennemført i foråret 2010. De nationale test var ét blandt flere elementer i et lovforslag fra december 2005 om fornyelse af folkeskolen for at forbedre det faglige niveau blandt eleverne gennem styrket, løbende evaluering i folkeskolen.

De nationale test tjener to formål: De skal fungere som et pædagogisk redskab til lærerne og har derudover et styringsformål rettet mod såvel institutioner, kommuner og på nationalt niveau (Undervisningsministeriet, 2005; 2006).

De nationale test tester eleverne i syv forskellige fag fra 2. til 8. klassetrin, heraf fire obligatoriske fag og tre frivillige fag. Samlet bliver det til 10 obligatoriske og op til 32 frivillige test i løbet af et skoleforløb.

Testene er it-baserede og adaptive, hvilket betyder, at de tilpasser sig den enkelte elev i sværhedsgrad i testforløbet. Hver test består af tre faglige profilområder og er selvscorende. Der gives tilbagemelding per profilområde samt en samlet vurdering. En test kan typisk gennemføres på én lektion, svarende til 45 minutter, med mulighed for at forlænge.

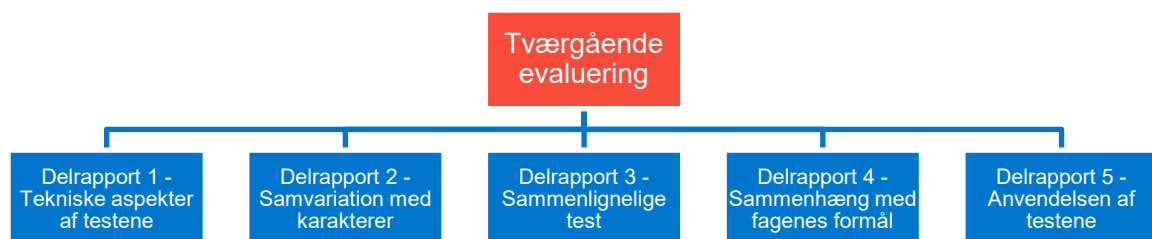
De nationale test blev senest evalueret i 2013. Det blev her besluttet, at der efter en femårig periode igangsættes en ny evaluering. VIVE udarbejder denne nye evaluering af de nationale test.

Evalueringen har til formål at belyse styrker såvel som svagheder omkring indholdet og brugen af de nationale test i folkeskolen samt give et vidensgrundlag, der kan danne afsæt for det fremadrettede arbejde med udvikling og brug af de nationale test i folkeskolen. Evalueringen svarer konkret på det følgende, overordnede evalueringsspørgsmål:

Evalueringsspørgsmål

Har de nationale tests indhold og udformning styrket skolernes evaluingskultur og derigennem elevernes faglige niveau?

Evalueringens underordnede undersøgelsesspørgsmål udmønter sig i seks rapporter og en bilagsrapport: én tværgående evalueringsrapport og fem delrapporter, der omhandler hvert sit emne samt en bilagsrapport til delrapport 5.



Hver delrapport besvarer selvstændige undersøgelsesspørgsmål under det overordnede spørgsmål. De fem delrapporter kan, ligesom den tværgående evaluering, læses selvstændigt.

De enkelte delrapporteringer besvarer følgende undersøgelsesspørgsmål:

Delrapport	Titel	Undersøgelsesspørgsmål
	Evaluering af de nationale test – tværgående rapport	Den tværgående rapport samler resultaterne fra de fem delrapporter
1	Review af evalueringen af de statistiske aspekter ved de nationale test	Har STIL på tilfredsstillende vis besvaret rådgivningsgruppens evalueringsspørgsmål om de nationale tests statistiske usikkerhed, reliabilitet og øvrige måleegenskaber?
2	De nationale tests samvariation med karakterer	Hvad er samvariationen mellem elevers præstationer i testene og karakterer i 8. og 9. klasses-prøverne?
3	Kortlægning af sammenlignelige test	Hvilke test findes, der i formål, indhold og omfang minder om de danske nationale test?
4	De nationale tests sammenhæng med fagenes formål	I hvilket omfang er der sammenhæng mellem de nationale test og de centrale dele af faget og fagenes formål jf. Fælles Mål?
5	Anvendelsen af de nationale test samt bilagsrapport med resultater fra spørgeskemaer	Hvordan opleves de nationale test som evalueringsredskab? Hvordan bruges de nationale test i dialogen og opfølgningen på tværs af lokale politikere, forvaltning, skoleledere, lærere, elever og forældre?

Denne rapport er Delrapport 2 og omhandler sammenhængen mellem resultater i de nationale test og folkeskolens 8. og 9. klasses-prøver. Rapporten viser, hvor stærk sammenhængen er mellem elevernes resultater i de nationale test og i folkeskolens 8. og 9. klasses-prøver. Rapporten skal ses i sammenhæng med Delrapport 1: Review af evalueringen af de statistiske aspekter ved de nationale test, der kritisk gennemgår de nationale test og vurderer testenes validitet og reliabilitet samt forslag til at forbedre disse; Delrapport 4, der ser på sammenhængen mellem opgaverne i de nationale test og målene for de konkrete fag, samt Delrapport 5, der behandler anvendelsen af resultaterne på både individ- og aggregeret niveau.

Sammenfatning

De nationale test har til formål at teste eleverne i centrale dele af fagene. Det vil sige, at de nationale test ikke tester eleverne i alt indhold i et fag, men kun områder af et fag. For at den viden, der skabes af de nationale test, bliver anvendt, kræves det, at testene opfattes som relevante i forhold til faget (Preskill & Boyle, 2008). Det skyldes, at eksempelvis lærere og skoleledere vil være mere tilbøjelige til at anvende de nationale test, hvis de kan se en umiddelbar fordel heri (Cousins & Shulha, 2006; Greene, 1988; Preskill, Zuckerman & Matthews, 2003).

Som en del af evalueringen af de nationale test har denne delrapport til formål at undersøge, i hvilket omfang der er sammenhæng mellem det faglige indhold, der testes på i de nationale test, og karakterer ved udskolingen. Dette gøres ved at gennemføre en empirisk analyse af samvariationen i elevers præstationer i de nationale test og karakterer i folkeskolens 8. og 9. klasses-prøver.

Konkret beskriver denne delrapport følgende:

Undersøgelsesspørgsmål

Hvad er sammenhængen mellem elevers præstationer i testene og karakterer?

Formålet med analysen er således at undersøge, hvorvidt de nationale test måler fagligt relevant – og dermed, hvor valide de nationale test er. Dette gøres ved at sammenholde elevernes resultater ved de nationale test med karakterer i 8. og 9. klasse. Delrapporten undersøger udelukkende de nationale tests kriterievaliditet, hvilket betyder, om de nationale test måler noget, der er relevant i forhold elevernes faglige præstationer inden for det givne fag. Det vil sige, at undersøgelsen kun forholder sig til testenes eksterne validitet. For en undersøgelse af de nationale tests pålidelighed eller reliabilitet henvises til Delrapport 1.

Resultater

Herunder præsenteres resultaterne af analysen af de nationale tests sammenhæng med karakterer.

Der er sammenhæng mellem resultater i de nationale test og 8. og 9. klasses-prøverne

Denne undersøgelse viser i lighed med andre, tidligere undersøgelser, at der er sammenhæng mellem de nationale test og senere karakterer i 8. klasses standpunktskarakterer eller folkeskolens 9. klasses-prøver. Sammenhængen mellem resultater i de nationale test og karakterer varierer mellem fagene, men er på niveau med sammenhænge mellem test og karakterer i andre lande.

Genberegning af de nationale tests sværhedsgrader giver ændringer i sammenhænge

Denne delrapport ser også på, hvorvidt de nationale test har mere eller mindre sammenhæng med folkeskolens 9. klasses-prøver, når der deles op på før og efter genberegningen af de nationale tests sværhedsgrader.

Disse genberegninger blev foretaget før skoleåret 2014/2015 for fagene, dansk (læsning), matematik, engelsk og biologi. I 2015/2016 blev de nationale tests sværhedsgrader genberegnet i fagene geografi og fysik/kemi. Generelt finder denne analyse, at ændringen har haft betydning for sammenhænge, men der er stor forskel mellem fag og inden for fag mellem fagets profilområder.

Der er stærke sammenhænge i dansk (læsning) og matematik og mindre stærke sammenhænge i fysik/kemi

Analysen viser, at særligt de nationale test i fysik/kemi samvarierer i mindre grad end øvrige fag med karakterer i 9. klasse i samme fag. Omvendt viser analyserne, at de stærkeste sammenhænge findes i fagene dansk (læsning) i 8. klasse og matematik i 6. klasse for de nationale test og karakterer i folkeskolens 9. klasses-prøver.

For fysik/kemi kan en medvirkende forklaring være, at fysik/kemi er et treårigt fag fra 7.-9. klasse, hvor læreren planlægger undervisningen af pensum på tværs af alle årene, men hvor den nationale test finder sted i 8. klasse. Det vil sige, at de nationale test potentielt indeholder emner, som eleverne ikke har gennemgået på testtidspunktet. Læs mere i Delrapport 4 om sammenhængen mellem opgaverne i testen og fagets mål.

Sammenhængen er på niveau med test i andre lande

Sammenlignet med andre analyser af standardiserede test og karakterer er de fundne sammenhænge på niveau med andre sammenlignelige analyser af standardiserede test. Og ud fra den betragtning er de nationale test på et forventeligt niveau. I denne undersøgelse af de danske nationale test og karakterer findes korrelationskoefficienter mellem 0,47 (fysik/kemi) og 0,74 (matematik). I amerikanske studier findes eksempler på korrelationskoefficienter på 0,31-0,66 mellem standardiserede test og karakterer. Og i et eksempel fra Sverige er korrelationskoefficienten fundet til at være 0,66-0,89 for sammenhænge mellem de svenske nationale test og matematikkarakterer.

Samling af profilområderne vil øge præcisionen i målingerne

Analysen viser, at et samlet mål for elevernes faglige præstationer for hver af de nationale test har større samvariation med karaktererne i folkeskolens 9. klasses-prøver i tilsvarende fag. Det vil sige, at mens resultater fra de enkelte profilområder har lavere præcision med hensyn til at forudsige elevernes præstationer i 9. klasse, så kan et samlet mål for hver national test være mere præcist med hensyn til at afdække elevernes faglige niveau.

Sammenhængen stiger med øget samtidighed

Generelt er sammenhængen støt stigende, jo tættere klassetrinet, hvor de nationale test tages, og karakterer i 9. klasse gives. Det vil sige, at øget samtidighed og sammenhæng korrelerer. Dette understøttes også af et yderligere fund, der viser sammenhængen mellem test taget i 8. klasse og standpunktskarakterer i 8. klasse, der er ligeså stærke eller stærkere end sammenhænge mellem test taget i 8. klasse og karakterer i 9. klasse.

Faglige resultater hænger ved

Analysen finder også, at elever, der havde lave resultater på de nationale test i tidligere klassetrin, også har lave karakterer i de samme fag ved folkeskolens 9. klasses-prøver. Tilsvarende viser analysen, at elever, der klarede sig godt på de nationale test på tidligere klassetrin, også får højere karakterer end de elever, der klarede sig mindre godt.

Undersøgelsen indikerer samlet, at der er en god sammenhæng mellem de nationale test og karakterer. Den indikerer endvidere, at de nationale test har en god ekstern validitet, om end de kun måler det, som er indeholdt i testen, og kun en andel af det, der undervises i.

Datagrundlag

Analysen er kvantitativ og baserer sig på registerdata fra Danmarks Statistik og STILs registre for de nationale test. Sammenhænge er undersøgt gennem korrelationer mellem elevernes faglige præstationer i de nationale test med karakterer i folkeskolens 9. klasses-prøver for de samme elever, der har taget de nationale test i dansk (2., 4., 6. og 8. klassetrin), engelsk (7. klassetrin), matematik (3. og 6. klassetrin), fysik/kemi, biologi og geografi i 8. klasse, samt standpunktskarakterer i 8. klasse for fagene dansk, matematik, fysik/kemi og engelsk. Oplysninger om de nationale test stammer fra skoleårene 2009/2010 til 2017/2018. Oplysninger om elevernes karakterer stammer fra skoleårene 2010/2011 til 2017/2018.

1 Indledning

Indførelsen af de nationale test i folkeskolen blev vedtaget af Folketinget i 2006 som led i en ændring i folkeskoleloven. De nationale test har til hensigt at styrke evalueringkulturen i folkeskolen og sikre en ensartet evaluering af elevernes faglige niveau på tværs af landet med det formål at forbedre det faglige niveau blandt eleverne (Børne- og Undervisningsministeriet, 2019).

Baggrunden for og indholdet af de nationale test beskrives i større detalje i den tværgående evalueringsrapport af de nationale test (Flarup, 2020).

1.1 Formål

Formålet med denne delrapport i evalueringen af de nationale test er at undersøge, i hvilket omfang der er sammenhæng mellem det faglige indhold, der testes på i de nationale test, og andre aspekter af fagenes formål, målt ved karakterer. Dette gøres ved at gennemføre en empirisk analyse af samvariationen i elevers præstationer i de nationale test og karakterer fra 8. og 9. klasses standpunktskarakterer og karakterer fra folkeskolens 9. klasses-prøver. Konkret bekræfter denne delrapport følgende:

Undersøgelsesspørgsmål

Hvad er sammenhængen mellem elevers præstationer i testene og karakterer?

Rapporten undersøger dermed de nationale tests eksterne validitet. Validiteten er i denne sammenhæng beregnet som en kriterievaliditet, hvor det undersøges, om de nationale test måler noget, der er relevant i forhold elevernes faglige præstationer.

Korrelationerne mellem de nationale test og karaktererne analyseres, mens de kausale sammenhænge mellem de nationale test og karakterer ikke undersøges. I stedet gives en deskriptiv analyse af samvariationen mellem de nationale test og karakterer fra 8. og 9. klasses standpunktskarakterer og karakterer fra folkeskolens 9. klasses-prøver.

1.2 Læsevejledning

Rapporten indledes med et kapitel om design og metode vedrørende de nationale test og måleforhold samt tidligere forskning på området. Dernæst gennemgås fagene enkeltvis i følgende rækkefølge: Dansk (læsning), Matematik, Biologi, Fysik/kemi, Geografi og Engelsk.

2 Design og metode

I det følgende beskrives design og metode til samvariationsanalysen mellem de nationale test og karakterer. Samvariationsanalysen undersøger de nationale tests prædikative egenskaber. Det vil sige, at det undersøges, hvorvidt der er sammenhæng mellem de nationale test på tidligere klassetrin og 9. klasses-prøverne samt 8. klasses standpunktskarakterer.

Analyserne i denne rapport beskæftiger sig udelukkende med de nationale tests kriterievaliditet. Det vil sige, hvordan de nationale test relaterer sig til andre test, men ikke, hvordan de enkelte dele af testen relaterer sig til sig selv (Pett, Lackey & Sullivan, 2003). Denne analyse forholder sig dermed ikke til den indre validitet af de nationale test, men derimod til den eksterne validitet af de nationale test, og om testene generelt måler noget, der kan relateres til elevernes faglige præstationer. Dette gøres ved at sammenligne de nationale test med karakterer i slutningen af folkeskolen. Som med en test-retest må det forventes, at jo mere samtidige test er, i dette tilfælde karakterer og de nationale test, jo større sammenhæng vil der være. Det betyder, at det forventes, at der er større sammenhænge mellem de nationale test i 8. klassetrin og 9. klasses karakterer, end der er mellem de nationale test i 2. eller 3. klasse.

En tests måleegenskaber deles op i to hovedkategorier: validitet og reliabilitet. Reliabiliteten er et mål for, hvor præcist en test måler det, der skal måles. Reliabilitet dækker bl.a. over to centrale aspekter ved en tests pålidelighed:

- Det første aspekt er den indre konsistens i en test. Det vil sige, hvor stærk sammenhængen er mellem de items, der skal måle en underliggende faktor såsom elevens evne til at læse.
- Det andet aspekt er stabilitet. En tests stabilitet måles ved en såkaldt test-retest. Dette er et mål for, om det blot er tilfældigheder, testen indfanger, eller om der er tale om systematik. Den underliggende antagelse er her, at personen, der vurderes med testen, ikke ændrer sin adfærd over tid. Det vil sige, at testen gives til de samme personer to eller flere gange, hvor der har været tid mellem målingerne.

Disse aspekter er klarlagt i både Bundsgaard & Kreiner (2019) og ministeriets egne afrapporteringer (Undervisningsministeriet, 2015).

I delrapportens analyser anvendes data fra skoleårene 2009/2010 til 2017/2018 fra de nationale test, og data fra skoleårene 2010/2011 til 2017/2018 med oplysninger om elevernes karakterer. I forhold til at anvende standpunktskaraktererne fra 8. klasse er der data for skoleåret 2017/2018, hvorfor der kun indgår en enkelt årgang i denne analyse, men flere i de øvrige analyser. Standpunktskaraktererne i 8. klasse dækker i data over følgende fag: dansk, matematik, engelsk og fysik/kemi. Data stammer fra Danmarks Statistiks registre over 9. klassesprøver i folkeskolen og STILs registre over resultater i de nationale test.

2.1 Rasch-modellen

Den bagvedliggende algoritme i de nationale test er baseret på en psykometrisk model kaldet Rasch-modellen.

Rasch-modellen er en statistisk model, der oprindeligt er udviklet til læseprøver og intelligens-test. I modellen bliver elevens dygtighed og opgavens sværhedsgrad målt på én og samme

skala. Elevens resultat (dygtighed) er defineret som sværhedsgraden på den opgave, hvor eleven har lige stor sandsynlighed for at svare rigtigt og forkert. Selve testen handler altså om at finde ud af, hvor svære opgaverne skal være, for at eleven har 50 % sandsynlighed for at svare rigtigt. Lidt forenklet kan man sige, at jo sværere en opgave er, desto færre elever kan svare på den – og omvendt. Man finder opgavernes sværhedsgrad ved at afprøve opgaverne empirisk. Rasch-modellens styrke er, at den testmetodisk sikrer test, hvor resultatet er et godt udtryk for elevernes samlede resultat, mens en ulempe er, at opgaverne tester et forholdsvis snævert fagligt område (Undervisningsministeriet, 2018).

Modellen har statistiske egenskaber, der gør det muligt at rangordne individer efter deres faglige niveau. Rangeringen af individer (i dette tilfælde elever) er baseret på en logit-skala, der kan være \pm uendelig, men i realiteten falder mellem -7 og 7. Eleven starter med en opgave med en sværhedsgrad midt i opgavebanken i hvert profilområde, og ændringerne i opgavesværhedsgrader er kun $\pm 0,5$ logit (tidligere ± 1 logit) og kun $\pm 0,25$ logit i enkelte profilområder.

I det følgende anvendes logit-skalaen fra Rasch-modellen til at analysere sammenhængen mellem de nationale test og karakterer ved 9. klasses-prøver. Logit-skalaen anvendes, da dette er den nemmest tolkelige del af skalaen for de nationale test. Det skyldes, at 1 logit er den samme, uanset hvor på logit-skalaen, man ser, mens de øvrige skalaer – norm- og kriteriebaserede skalaer – ikke på samme måde har samme tolkning. Til at sammenligne testene følges metoden beskrevet i Beuchert & Nandrup (2018), hvor hvert profilområde først standardiseres, så der er en middelværdi på 0 og en standardafvigelse på 1. Dette gøres for hvert fag, profilområde og år. Det sikrer, at der kan sammenlignes på tværs af fag og profilområder.

Dernæst tages gennemsnittet for hver enkelt elev af de tre profilområder, som et fag udgøres af. Dette gennemsnit bliver også standardiseret for hvert fag og år, så der er en middelværdi på 0 og en standardafvigelse på 1. Det gør det muligt at sammenligne resultaterne som standardafvigelser og letter dermed tolkningen af resultaterne.

Der anvendes udelukkende resultater fra de obligatoriske nationale test. Der er mål for hver af de tre profilområder under hvert fag, hvilket også samles til ét mål for hver af de nationale test. Dette giver en mere jævn fordeling og giver et samlet billede af elevernes faglige præstationer inden for det specifikke fag. I matematik i 3. klasse er profilområdet "Matematik i anvendelse" siden 2014 blevet udskiftet med profilområdet "Statistik og sandsynlighed". Tilsvarende ændring blev i 2017 foretaget i matematik i 6. klasse. Dette tages der højde for i analyserne ved, at der ses på korrelationer før og efter ændring af fagenes sværhedsgrad, som ligger på hver sin side af denne ændring i profilområdet, og ved standardiseringen over år og profilområder. Som afhængig variabel anvendes karakterer ved folkeskolens 9. klasses-prøve. Der anvendes dansk-karakterer i læsning, skriftlig fremstilling, retskrivning og mundtlig eksamen – både hver for sig og i gennemsnit. Der bruges også karakterer i matematik (med og uden hjælpemidler), engelsk (mundtlig), biologi, fysik/kemi, biologi og geografi. En ændring i eksamensformen i fagene fysik/kemi, biologi og geografi betyder, at disse fag ikke anvendes som udfaldsmål fra skoleåret 2016/2017, da de er blevet samlet i en fælles prøve. Det gør det vanskeligt at adskille den reelle sammenhæng mellem de nationale test i faget og karaktererne i folkeskolens 9. klasses-prøve for disse tre fag.

For nogle fag gør det sig gældende, at sværhedsgraderne er blevet genberegnet i skoleåret 2014/2015. Det vil sige, at resultater fra de nationale test ikke er direkte sammenlignelige fra skoleårene før 2014/2015. Denne ændring gør sig gældende for fagene dansk, matematik, engelsk og biologi. For fagene fysik/kemi og geografi blev sværhedsgraderne genberegnet i

skoleåret 2015/2016. For fysik/kemi og geografi er det derfor ikke direkte muligt at sammenligne resultater før skoleåret 2015/2016 med resultater efter denne genberegning af sværhedsgrader. Oplysningerne om disse ændringer er fundet i "National præstationsprofil – de nationale test"^{1,2}.

I Tabel 2.1 fremgår de mulige koblinger mellem de nationale test og karakterer i 9. klasse. Som det fremgår af tabellen, er der forskel på, hvornår de nationale test er aflagt, og hvornår eleven tager afgangsprøve i 9. klasse. Tabellen viser, at der går 7 år, fra prøven i de nationale test i dansk (læsning) 2. klasse er aflagt, til det er muligt at se på sammenhængen mellem denne og karaktererne for folkeskolens 9. klasses-prøve. Derfor er der kun 2 år, hvor sammenhængen mellem dansk i 9. klasse og de nationale test i 2. klasse kan ses. Tilsvarende gør sig gældende for de øvrige fag. Fagene i 8. klasse er dog kun forsinket med et enkelt år i forhold til karaktererne i tilsvarende fag.

Fagene fysik/kemi, biologi og geografi indgår i en samlet fællesprøve fra skoleåret 2016/2017 i 9. klasse. Det kan dermed være vanskeligt at udlede, hvilke fag der er testet og fylder mest i denne sammenligning. Derfor undersøges sammenhængen mellem de nationale test i disse fag og de tilsvarende fag frem til og med skoleåret 2015/2016. For både biologi og geografi gør det sig gældende, at de er udtræksfag i folkeskolens 9. klasses-prøve. Det betyder, at færre elever har aflagt prøve i disse fag end i eksempelvis dansk og matematik, hvilket kan betyde, at der er større usikkerhed om sammenhængen mellem disse fag.

Tabel 2.1 Tabel over undersøgte sammenhænge mellem 9. klasses karakterer og de nationale test i de enkelte fag, delt på skoleår.

Fag i nationale test	År for karakter i folkeskolens 9. klasseprøve								
	2009/ 2010	2010/ 2011	2011/ 2012	2012/ 2013	2013/ 2014	2014/ 2015	2015/ 2016	2016/ 2017	2017/ 2018
Dansk (læsning) 2. klasse								x	x
Dansk (læsning) 4. klasse						x	x	x	x
Dansk (læsning) 6. klasse					x	x	x	x	x
Dansk (læsning) 8. klasse		x	x	x	x	x	x	x	x
Matematik 3. klasse							x	x	x
Matematik 6. klasse				x	x	x	x	x	x
Engelsk 7. klasse			x	x	x	x	x	x	x
Fysik/kemi 8. klasse		x	x	x	x	x	x	(x)	(x)
Biologi 8. klasse (udtræksfag i 9. klasse)		x	x	x	x	x	x	(x)	(x)
Geografi 8. klasse (udtræksfag i 9. klasse)		x	x	x	x	x	x	(x)	(x)

Anm.: x angiver, at der er muligt at se på sammenhængen mellem faget, som de nationale test undersøger, og et tilsvarende fag ved folkeskolens 9. klasses-prøve.

(x) angiver, at faget ikke er fuldt sammenligneligt med senere år, da 9. klasses-eksamen er blevet ændret.

Note: Fra skoleåret 2016/2017 er fysik/kemi, biologi og geografi samlet i en fællesprøve i folkeskolens 9. klasses-prøve.

¹ <https://www.uvm.dk/statistik/grundskolen/karakterer-og-test/nationale-test/national-praestationsprofil> tilgået 29/10-2019.

² Der er endvidere testet, om der er statistiske forskelle på, hvorvidt opdelingen før og efter dette databrud har betydning i forhold til at undersøge disse forhold uden denne opdeling. I brede træk er der statistiske forskelle, uden disse er substantielt større end de opdelinger, der her er foretaget. Disse opdelinger beholdes, da STIL i deres dokumentation også gør opmærksom på, at resultater før og efter denne opdeling ikke er sammenlignelige. Tilsvarende viser Bundsgaard & Kreiner (2019), at der for dansk (læsning) i 8. klasse kan være vanskeligheder med at sammenligne de nationale test før og efter ændringen af sværhedsgraderne i de nationale test.

2.1.1 Ændring af de nationale test

I 2014 foretog Undervisningsministeriet en genberegning af sværhedsgraderne af opgaverne i de nationale tests opgavebanker (Bundsgaard & Kreiner, 2019). Tilsvarende blev der foretaget ændringer i algoritmen, så der for de nationale test fra 2014/2015 og fremefter var tale om, at testen havde tre opgaver til at finde den rette sværhedsgrad.

Ændringerne i algoritmen blev foretaget forud for de frivillige test i efteråret 2014. Ændringerne bestod i, at run-in-perioden, hvor de nationale test søger at beregne et udgangspunkt for elevdygtigheden, afkortes fra fem til tre opgaver. Endvidere bestod denne ændring i, at eleven starter med en opgave med en sværhedsgrad midt i opgavebanken i hvert profilområde, og at ændringerne i opgavens sværhedsgrader kun er $\pm 0,5$ logit (tidligere ± 1 logit) og kun $\pm 0,25$ logit i enkelte profilområder. Denne ændring i algoritmen bag de nationale test omfattede alle test på samme tidspunkt (Undervisningsministeriet, 2015). Ændringen af sværhedsgrader betyder, at der kan være statistisk støj, som kan påvirke sammenhængen mellem de nationale test og karaktererne fra folkeskolens 9. klasses-prøve.

2.2 Metode

Analysen fokuserer på sammenhængen mellem de nationale test og karakterer i det tilsvarende fag i 8. klasses standpunktskarakterer og 9. klasses-prøven med udgangspunkt i lineære sammenhænge. Det vil sige, at eksempelvis sammenhængen mellem de nationale test i dansk (læsning) og karakterer i folkeskolens 9. klasses-prøver i dansk kan beskrives sådan, at for hver gang præstationen i den nationale test stiger med én enhed (logit eller standardafvigelse), så er betydningen den samme på karaktererne, uanset hvor i fordelingen af de nationale test eller karakterer eleven befinder sig.

Der anvendes korrelationer til at undersøge sammenhængen mellem de nationale test og karakterer. Korrelationer er den simple sammenhæng mellem to variable, som i denne sammenhæng er resultater i de nationale test og det tilsvarende fag ved folkeskolens 9. klasses-prøve. I forhold til korrelationer, så kan de falde mellem -1 og $+1$ og er udtryk for den sammenhæng, der – i dette tilfælde – er mellem to test.

I det følgende anvendes to typer af analyser, korrelationer og gennemsnit opdelt på elevers præstationer på de nationale test.

Først anvendes korrelationerne imellem de nationale test og karakterer. Disse korrelationer viser de lineære sammenhænge imellem de nationale test og karakterer. Analyserne viser ikke, hvordan grupper af elever, der eksempelvis klarer sig meget dårligt eller godt på de nationale test, sidenhen opnår karakterer ved folkeskolens 9. klasses-prøver.

Dernæst opdeles eleverne i grupper, der baserer sig på elevernes præstationer ved de nationale test, for at give en analyse af gruppesammenhænge mellem elevernes resultater på de nationale test og karakterer. Grupperne er eksempelvis defineret ved de 0-10 % lavest scorende ved de nationale test i et givet år. Tilsvarende inddeling laves for de elever, der var blandt de 11-20 % lavest scorende ved de nationale test, de 21-30 % lavest scorende osv. op til de 90-100 % højest scorende elever på de nationale test et givet år. For alle disse grupper af elever beregnes gennemsnit af karakterer i det fag, som de nationale test sammenholdes med. Gennemsnittet indsættes i en figur, hvor der er opdelt i perioder, før og efter de nationale tests

sværhedsgrader er blevet genberegnet, hvis det muligt for den pågældende test. I STILs dokumentation af de nationale tests reliabilitet fremgår der også sammenhænge imellem de nationale test målt ved de forskellige afrapporteringsformer (på henholdsvis normbaseret og kriteriebaseret skala), og her er der lineær sammenhæng mellem de nationale test i dansk (læsning) og karakter ved afgangsprøven i 9. klasse. Tilsvarende findes for matematik (Ministeriet for Børn, Undervisning og Ligestilling, 2016b). Den lineære sammenhæng er her, at elever, der klarede sig mindre godt på de nationale test i 8. klasse, også havde lavere karakterer i det tilsvarende fag i 9. klasse. Da analysen i denne rapport anvender data fra samme kilde, forventes det også, at der er tilsvarende sammenhænge.

Korrelationerne, der anvendes i analyserne, er baseret på Spearman's Rank Order-korrelation, der er mindre følsom over for ekstreme observationer som eksempelvis elever, der har scoret meget højt eller meget lavt på de nationale test end den normalt brugte Pearson-korrelation³. I forbindelse med analyserne har Fisher's z-transformation også været anvendt til at give en vurdering af usikkerheden af korrelationerne. På grund af de store antal af observationer giver Fisher's z-transformation nogle meget snævre konfidensbånd, der ligger meget tæt på Pearson-korrelationerne. Da disse konfidensbånd ikke har givet yderligere information, end hvad der kunne opnås igennem Pearson-korrelationer, er de derfor udeladt af denne delrapport.

Til at teste om der er forskelle mellem de korrelationer, der er før og efter en genberegning af de nationale tests sværhedsgrader, anvendes Jennrich's (1970) Wald test, som tester, om korrelationskoefficienterne fra to test er lige. Forventningen er, at hvis der er betydelige forskelle som følge af genberegningen af de nationale tests sværhedsgrader, så vil dette komme til udtryk i de sammenhænge, der er mellem de nationale test og karakterer før og efter denne ændring. Hvis ændringen er af stor betydning, så vil der være markante ændringer i sammenhængene.

Der er foretaget regressionsanalyser, som tager højde for elevernes sociale baggrundsforhold og for at skabe en mere sammenlignelig gruppe af elever før og efter genberegningen af sværhedsgrader. Disse analyser viser i det store hele det samme som det, der findes ved de simple korrelationsanalyser præsenteret i denne rapport. Da disse analyser ikke bringer yderligere information, er de ikke præsenteret i denne rapport.

De rapporterede resultater er statistisk signifikante på mindst et 5 %-niveau. Det vil sige, at man betragter det som usandsynligt, at sammenhænge mellem elevernes faglige præstationer i de nationale test og karakterer i 9. klasses afgangsprøve fremkommer på grund af statistiske tilfældigheder, idet sandsynligheden herfor er på under 5 %.

I forlængelse heraf er det vigtigt at understrege, at selvom det er statistisk signifikante forskelle, så er det ikke ensbetydende med, at forskellen nødvendigvis har substantiel betydning. At et resultat er statistisk signifikant, betyder ikke i sig selv, at det er interessant eller relevant i praksis. I det følgende gennemgås resultaterne for sammenhængene mellem de nationale test og karakterer.

³ Korrelationerne er denne forbindelse også beregnet som Pearson-korrelationer, og resultaterne fra disse analyser er meget lig dem, der bliver præsenteret i denne rapport.

2.3 Tidligere forskning

Forskning har vist, at de nationale test kan bruges til at forudsige senere placeringer på de nationale test. For eksempel viser en analyse fra Rockwool Fondens Forskningsenhed, at en elevs resultat ved de nationale test i dansk (læsning) i 2. klasse kan bruges til at forudsige elevens resultat ved 8. klasses nationale test i dansk (læsning) (ROCKWOOL Fonden, 2019). I et andet studie har forskere vist, at der er sammenhæng mellem tidligere tagne test og resultater på senere test (Calmar Andersen & Nielsen, 2019). I en gennemgang af de nationale test viser forskere, at der er forskel på test, der er taget før og efter en ændring af de nationale tests sværhedsgrader i 2014 (Bundsgaard & Kreiner, 2019).

Den tidligere forskning i sammenhængen mellem de nationale test og andre test, herunder karakterer, har fundet, at der er sammenhæng mellem de nationale test og andre test. Forskere har således set på den prædiktive validitet af de nationale test på karakterer fra folkeskolens 9. klasses-prøver. Resultaterne fra disse har vist, at de nationale test i dansk er højt associeret med 9. klasses-prøver i både dansk, matematik og fysik/kemi. Størst sammenhæng er der dog for de fag, som de nationale test ligger inden for, hvilket eksempelvis vil sige, at sammenhængen er størst mellem de nationale test i fysik/kemi og afgangsprøve i fysik/kemi og tilsvarende for engelsk, dansk og matematik (Beuchert & Nandrup, 2018). Børne- og Undervisningsministeriet har i forbindelse med dokumentationen af de nationale test også undersøgt sammenhængen mellem de nationale test og karakter i folkeskolens 9. klasses-prøver. Disse analyser begrænser sig til et enkelt år og samt til dansk (læsning) i 8. klasse og matematik i 6. klasse (Ministeriet for Børn, Undervisning og Ligestilling, 2016a).

Trods sammenhæng er det centralt at pointere, at karakterer og nationale test ikke forventes at måle præcis det samme, hvorfor der ikke forventes perfekte sammenhænge (dvs. korrelationskoefficienter på 1), når disse sammenlignes. Typisk anses en sammenhæng på korrelationskoefficient på 0,7 som udtryk for en høj sammenhæng mellem to test, da det betyder, at ca. 50 % af variationen af den ene test kan forklares med den anden (se eksempelvis Nunnally & Bernstein, 1994; Pett, Lackey & Sullivan, 2003). I forhold til lignende test i Sverige er der en sammenhæng mellem nationale test og karakterer i slutningen af grundskolen på 0,66 til 0,89 for matematik i Sverige (Wiberg, 2019), hvor de mest samtidige test og karakterer har de højeste sammenhænge. Hvis man ser mere generelt på sammenhængen imellem standardiserede test, som de nationale test, og karakterer, så viser et review af karakterer og standardiserede test, at man typisk finder sammenhænge på 0,31 til 0,66 (Brookhart et al., 2016). I denne forstand er det ikke forventeligt at finde korrelationer, der er over 0,7. Der tages derfor udgangspunkt i, at de forventede sammenhænge mellem de nationale test og karakterer skal være mellem 0,3 og 0,7. En korrelationskoefficient betragtes dermed som god, jo tættere den er på 0,7 – og gerne over 0,7.

Det forventes endvidere, at sammenhængen mellem de nationale test og karakterer vil være større, jo mere samtidige testene er. Eksempelvis forventes sammenhængen mellem de nationale test i 8. klasses-trin og karakterer i 9. klasse at være større, end den vil være mellem de nationale test i 2. klasses-trin og karakterer i 9. klasse. Sammenhængen mellem de nationale test i 8. klasse og standpunktskaraktererne i 8. klasse vises i analysen, for at vise den størst mulige samtidighed i mellem de nationale test og karakterer. Det giver også et udtryk for den størst muligt forventede korrelation imellem disse.

2.3.1 Muligheder for målefejl i test

Som ved test generelt er der tale om, at de nationale test har målefejl. Forskning har også vist, at tidspunktet på dagen, hvor eleverne tager de nationale test, kan have betydning for deres resultat i testen (Sivertsen, Gino & Piovesan, 2016). Ligeledes viser forskningen, at lærernes undervisningskompetencer kan påvirke elevernes resultater fra de nationale test (Kristensen & Skov, 2019). Alt i alt betyder disse forhold, at de nationale test ikke er et fuldstændigt mål for elevernes færdigheder eller kompetencer. Ligeledes påvirkes elevernes karakterer af forskellige ydre forhold såsom elevsammensætninger og lærernes undervisningskompetencer, ligesom de også kan være påvirket af indre forhold såsom motivation, kognitive færdigheder og elevernes sociale baggrund. Dette betyder, at hverken karakterer eller de nationale test er fejlfrie i den måde, hvorpå elevernes færdigheder måles. Her skal det også bemærkes, at sammenhængen mellem de nationale test og karakterer fra folkeskolens 9. klasses-prøver kan være påvirket af andre ting, fx folkeskolereformen og at eleverne over tid lærer at tage de nationale test.

Derudover er der en række ting, der gør, at karakterer og resultater fra de nationale test ikke nødvendigvis måler det samme. Blandt andet indeholder de nationale test elementer, der relaterer sig til tre specifikke profilområder, som er relateret til Fælles Mål, mens karakterer også måler andre elementer, der er relateret til faget. Alt i alt betyder disse forhold, at der ikke forventes et ligeligt forhold mellem de nationale test og karakterer i 9. klasse. Der er dog tilpas stort overlap mellem de nationale test og karakterer i folkeskolens 9. klasses-prøver til, at en sammenligning af disse er meningsgivende. VIVEs delrapport 4, "De nationale tests sammenhæng med fagenes formål", fra denne evaluering, uddyber pointen ved en kortlægning og analyse af sammenhængen mellem indholdet af de nationale test og de enkelte fags Fælles Mål. Delrapport 4 afdækker overlappet mellem de konkrete opgaver i de nationale test og de opstillede Fælles Mål inden for de fire obligatoriske testfag dansk, matematik, engelsk og fysik/kemi. Delrapporten viser også, at de nationale test i dansk og engelsk måler relativt snævert i forhold til Fælles Mål, mens de nationale test i matematik og fysik/kemi måler væsentligt bredere. Det betyder, at man ikke kan sammenligne direkte imellem fag, men kun inden for fagene.

2.3.2 Sammenhængen mellem de nationale test og PISA

Sammenhængen mellem de nationale test og PISA-testene fra 2012 også blevet undersøgt. Her er der også fundet sammenhænge mellem elevernes PISA-score i læsning og matematik og de nationale test i dansk (læsning) og matematik (DAMVAD, 2014). Sammenhængen mellem præstationer i de nationale test og PISA gør sig også gældende, når der tages højde for elevernes socioøkonomiske baggrund, køn og etnicitet. De socioøkonomiske baggrundsvariable viser, at elever, der kommer fra mere ressourcerstærke hjem, har en højere score i både de nationale test og PISA-scorer. Resultaterne viser, at de nationale test kan bruges prædiktivt i den forstand, at en elev, der scorer (lavt) højt på de nationale test i et fag, vil have større sandsynlighed for at opnå en (lav) høj PISA-score i det tilsvarende fag. Det gør sig gældende både for sammenhængen mellem de nationale test og PISA-testene, og når der tages højde for socioøkonomisk baggrund (ibid.).

3 Dansk (læsning)

I det følgende vises resultaterne mellem de nationale test i dansk (læsning), de tre profilområder og karaktererne i 9. klasse dansk, herunder læsning, skriftlig fremstilling, retskrivning og mundtlig eksamen og gennemsnittet af disse karakterer.

Gennemgangen af resultaterne i dansk (læsning), indeholder en kort beskrivelse af, hvordan resultaterne skal læses, mens gennemgangen af de øvrige nationale test kun viser resultaterne uden yderligere beskrivelse. Gennemgangen af dansk (læsning) skal dermed ses som en guide til, hvordan de tilsvarende tabeller for de øvrige fag kan læses og tolkes.

3.1 Dansk (læsning) i 2. klasse

For dansk (læsning) i 2. klasse er der endnu ingen, der på nuværende tidspunkt har deltaget i de nationale test og 9. classes-prøver, efter sværhedsgraderne i dansk (læsning) er blevet ændret. Derfor dækker tabellerne kun test taget i perioden før ændringen af de nationale tests sværhedsgrader i dansk, hvilket vil sige før skoleåret 2014/2015 for dansk (læsning). Tabel 3.1 viser korrelationerne mellem de nationale test i dansk (læsning) ved 2. klassetrin og karakterer i folkeskolens 9. classes-prøver i dansk.

Tabellen viser for det første, at der er statistisk sikre sammenhænge mellem de nationale test i 2. klasse og karakterer i 9. klasse, og at sammenhængen er størst mellem det samlede mål for dansk (læsning) i 2. klasse og de samlede karakterer i folkeskolens 9. classes-prøve i dansk (0,594). Det samlede mål for dansk (læsning) i 2. klasse har nemlig lidt højere sammenhæng med folkeskolens 9. classes-prøver i dansk end de enkelte profilområder, der har henholdsvis 0,573; 0,563 og 0,419 for "Tekstforståelse", "Afkodning" og "Sprogforståelse".

For det andet ses det, at der er store forskelle mellem profilområderne og de dele af danskfaget, som afgangsprøven i 9. klasse består af. Her har profilområderne "Tekstforståelse" og "Afkodning" de højeste sammenhænge med karakter i læsning og retskrivning, mens profilområdet "Sprogforståelse" har lavere korrelationer.

For det tredje ses det, at de skriftlige og mundtlige eksamener i dansk i 9. klasse har de laveste sammenhænge med de nationale test i 9. klasse. Dette kan hænge sammen med, at den nationale test i dansk (læsning) er en målrettet læsetest, hvilket ikke nødvendigvis giver sig til udtryk i de skriftlige og mundtlige områder af danskprøverne i 9. klasse.

Tabel 3.1 Korrelationer mellem de nationale test i dansk (læsning) ved 2. klassetrin og karakterer i folkeskolens 9. classes-prøver i dansk

Karakterer	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse
Samlet dansk, 9. klasse	0,594***	0,573***	0,563***	0,419***
Læsning, 9. klasse	0,541***	0,527***	0,489***	0,397***
Mundtlig, 9. klasse	0,361***	0,348***	0,328***	0,273***
Retskrivning, 9. klasse	0,599***	0,573***	0,605***	0,387***
Skriftlig, 9. klasse	0,412***	0,395***	0,394***	0,290***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* p < 0.05, ** p < 0.01, *** p < 0.001. Antallet af observationer er 96.880.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

3.2 Dansk (læsning) i 4. klasse

I det følgende præsenteres resultaterne for dansk (læsning) i 4. klasse i forhold til folkeskolens 9. klasses-prøve. Af Tabel 3.2 fremgår korrelationer mellem de nationale test i dansk (læsning) ved 4. klassetrin og karakterer i folkeskolens 9. klasses-prøver i dansk. Korrelationerne mellem resultater ved 4. klasses nationale test i dansk (læsning) og karakter ved folkeskolens 9. klasses-prøver er højere end dem fundet ved 2. klasses nationale test. Det er dog ikke overraskende, givet at testene rent tidsmæssigt er tættere på hinanden, og dermed er graden af samtidighed også større. Det bemærkes dog, at profilområdet "Sprogforståelse" også her generelt er lavere end de øvrige profilområder. Her finder analysen, at sammenhængen mellem de nationale test i dansk (læsning) ved 4. klassetrin og 9. klasses karakterer er større for det samlede mål for elevernes faglige præstationer ved den nationale test end for de enkelte profilområder.

Tabel 3.2 Korrelationer mellem de nationale test i dansk (læsning) ved 4. klassetrin og karakterer i folkeskolens 9. klasses-prøver i dansk

Karakterer	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse
Samlet dansk, 9. klasse	0,664***	0,608***	0,590***	0,539***
Læsning, 9. klasse	0,611***	0,577***	0,520***	0,504***
Mundtlig, 9. klasse	0,404***	0,385***	0,327***	0,346***
Retskrivning, 9. klasse	0,657***	0,569***	0,639***	0,507***
Skriftlig, 9. klasse	0,477***	0,433***	0,429***	0,386***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.

* p < 0.05, ** p < 0.01, *** p < 0.001. Antallet af observationer er 187.082.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

3.3 Dansk (læsning) i 6. klasse

For dansk (læsning) i 6. klasse er det muligt at dele beregningen op i perioden før og efter genberegningen af de nationale test. Tabel 3.3 viser korrelationer mellem de nationale test i dansk (læsning) ved 6. klassetrin og karakterer i folkeskolens 9. klasses-prøver i dansk.

Det fremgår af Tabel 3.3, at korrelationerne mellem de nationale test i dansk (læsning) i 6. klasse og 9. klasses-prøver er højere, end de er for de nationale test i 2. og 4. klasse. Det betyder, at resultaterne fra 6. klasses nationale test i dansk er lidt bedre til at forudsige karaktererne i 9. klasse, end de er for især 2. klasse, og understøtter dermed hypotesen om, at sammenhængen mellem de nationale test og karakterer bliver større, jo mere samtidige målene er. De samlede mål for dansk (læsning) i 6. klassetrin har den største sammenhæng med folkeskolens 9. klasses-prøver af alle de sammenhænge, der afrapporteres i tabellen. Dette gælder både før og efter genberegningen af de nationale test. Tabellen viser dog, at der overordnet set er et fald i korrelationernes størrelse, når man sammenligner før og efter genberegning af sværhedsgraden for de nationale test i dansk (læsning). Eksempelvis er sammenhængen mellem profilområdet "Afkodning" og læsning i 9. klasse gået fra 0,565 til 0,547. Samtidig er sammenhængen mellem profilområdet "Tekstforståelse" og læsning i 9. klasse gået fra 0,583 til 0,596. Det betyder, at der er nogle ændrede sammenhænge, som kan medføre både lidt mindre og lidt større sammenhæng. Overordnet set er der tale om mindre ændringer fra før ændringen af sværhedsgraderne til efter.

Tabel 3.3 Korrelationer mellem de nationale test i dansk (læsning) ved 6. klassetrin og karakterer i folkeskolens 9. klasses-prøver i dansk. Før og efter genberegning af sværhedsgrad.

Karakterer	Nationale test før skoleåret 2014/2015				Nationale test efter skoleåret 2013/2014			
	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse
Samlet dansk, 9. klasse	0,692***	0,627***	0,652***	0,532***	0,679***	0,626***	0,620***	0,497***
Læsning, 9. klasse	0,629***	0,583***	0,565***	0,497***	0,626***	0,596***	0,547***	0,476***
Mundtlig, 9. klasse	0,438***	0,416***	0,380***	0,350***	0,429***	0,412***	0,359***	0,326***
Retskrivning, 9. klasse	0,665***	0,571***	0,687***	0,490***	0,664***	0,581***	0,663***	0,464***
Skriftlig, 9. klasse	0,540***	0,485***	0,514***	0,411***	0,461***	0,421***	0,432***	0,333***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer. * p < 0.05, ** p < 0.01, *** p < 0.001. Antallet af observationer er 205.268 før skoleåret 2014/2015 og 101.095 efter skoleåret 2013/2014. Der er statistisk signifikante forskelle på profilområderne og det samlede mål for de nationale test i dansk (læsning) og karaktererne i dansk i 6. klasse før og efter genberegning af fagets sværhedsgrader. Der er dog ikke statistisk signifikante forskelle på den samlede dansk karakter og tekstforståelse samt det samlede mål for de nationale test i dansk (læsning) i 6. klasse samt retskrivning.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

3.4 Dansk (læsning) i 8. klasse

De følgende tabeller viser sammenhængen mellem de nationale test i dansk (læsning) ved 8. klassetrin og karaktererne fra folkeskolens 9. klasses-prøve i dansk, opdelt på før og efter genberegningen af de nationale test samt standpunktskarakterer i 8. klasse.

Tabel 3.4 viser korrelationer mellem de nationale test i dansk (læsning) ved 8. klassetrin og karakterer i folkeskolens 9. klasses-prøver i dansk. Denne tabel viser korrelationerne opdelt på perioden før og efter genberegning af sværhedsgraden af de nationale test i dansk (læsning). Som det fremgår af tabellen, er sammenhængene mellem 9. klasses karakter i dansk og de nationale test generelt høje, når der sammenlignes med de sammenhænge, der findes i andre lande (se tidligere gennemgang). Det bemærkes dog, at korrelationerne mellem profilområderne "Afkodning" og "Sprogforståelse" og karaktererne i folkeskolens 9. klasses-prøve er væsentligt lavere efter genberegningen af de nationale test, end de er før. Det betyder overordnet set, at det for disse profilområder er sværere at forudsige de karakter, som eleverne opnår i 9. klasse på disse profilområder. De samlede mål for dansk (læsning) i 8. klassetrin har dog større sammenhæng med folkeskolens 9. klasses-prøver. Dette gælder både før og efter genberegningen af de nationale test. Tabellen viser også sammenhængen mellem de nationale test i 8. klasse og standpunktskaraktererne i 8. klasse. Disse sammenhænge er generelt på samme niveau, som de er for folkeskolens 9. klasses-prøver med den undtagelse, at der for profilområdet "Afkodning" generelt er større sammenhænge med de nationale test i 8. klasses standpunktskarakterer, end der er mellem folkeskolens 9. klasses-prøver og dette profilområde. Det vil sige, at de nationale test i dansk (læsning) har de samme prognostiske egenskaber, om der ses på 8. eller 9. klasse. Samtidigheden mellem karakterer og de nationale test i 8. klasse, der må anses for at være den højest mulige sammenhæng, betyder, at sammenhængen mellem de nationale test i 8. klasse og karakterer fra 9. klasses-prøver i dansk er på et forventeligt niveau.

Tabel 3.4 Korrelationer mellem de nationale test i dansk (læsning) ved 8. klassetrin og karakterer i folkeskolens 9. klasses-prøver i dansk. Før og efter genberegning.

Karakterer	Nationale test før skoleåret 2014/2015				Nationale test efter skoleåret 2013/2014			
	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse
Samlet dansk, 9. klasse	0,706***	0,668***	0,619***	0,526***	0,691***	0,670***	0,488***	0,398***
Læsning, 9. klasse	0,647***	0,620***	0,536***	0,500***	0,612***	0,616***	0,397***	0,357***
Mundtlig, 9. klasse	0,465***	0,463***	0,369***	0,358***	0,453***	0,459***	0,291***	0,257***
Retskrivning, 9. klasse	0,681***	0,606***	0,650***	0,485***	0,670***	0,607***	0,531***	0,388***
Skriftlig, 9. klasse	0,571***	0,541***	0,500***	0,420***	0,500***	0,478***	0,368***	0,292***
Samlet dansk, 8. klasse	-	-	-	-	0,676***	0,650***	0,629***	0,392***
Læsning, 8. klasse	-	-	-	-	0,638***	0,618***	0,567***	0,391***
Mundtlig, 8. klasse	-	-	-	-	0,506***	0,510***	0,452***	0,298***
Retskrivning, 8. klasse	-	-	-	-	0,638***	0,575***	0,633***	0,365***
Skriftlig, 8. klasse	-	-	-	-	0,550***	0,536***	0,521***	0,308***

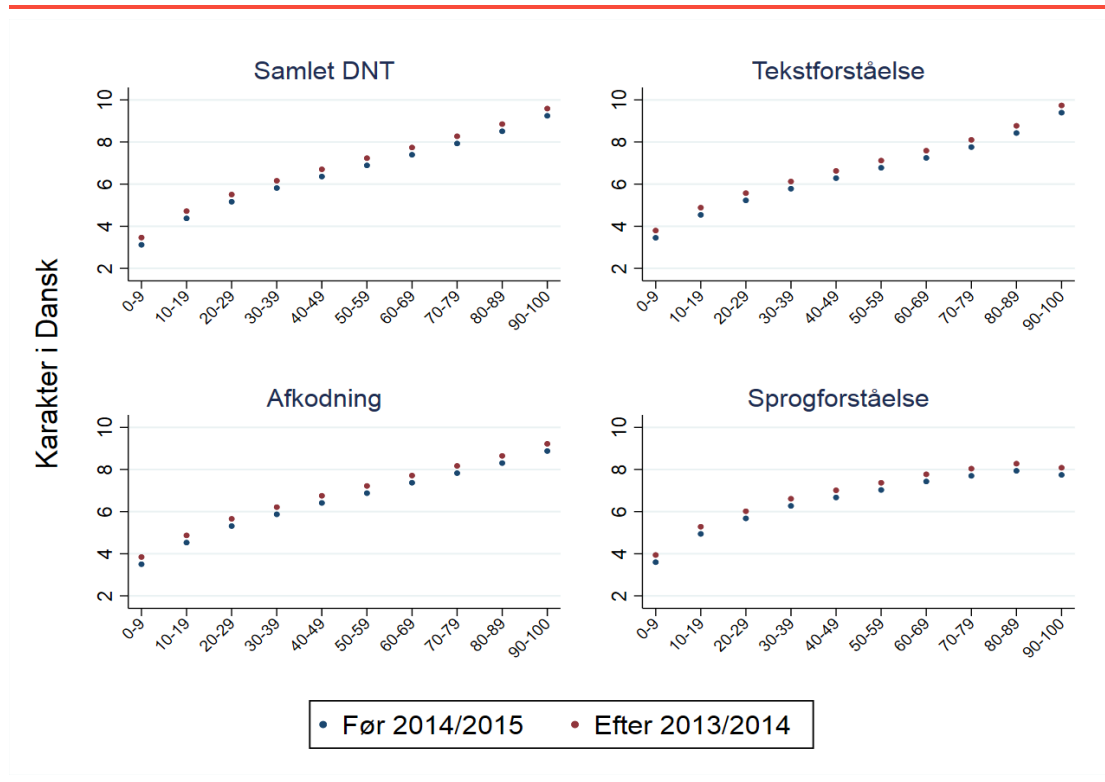
Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer. * p < 0.05, ** p < 0.01, *** p < 0.001. Antallet af observationer er 192.746 før skoleåret 2014/2015 og 193.716 efter skoleåret 2013/2014. For sammenhænge med standpunktskarakterer i 8. klasse er antallet af observationer 48.942. Der er statistisk signifikante forskelle på profilområderne og de samlede mål for de nationale test i dansk (læsning) og karaktererne i dansk i 9. klasse før og efter genberegning af fagets sværhedsgrader. Der er dog ikke statistisk signifikante forskelle på den samlede dansk karakter og tekstforståelse og mundtlig eksamen i 9. klasse og profilområdet "Tekstforståelse" samt det samlede mål for de nationale test i dansk (læsning) i 9. klasse og retskrivning.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Sammenligner man resultaterne med resultaterne i Tabel 3.3 (for 6. klassetrin), så fremgår det, at der overordnet set er lidt større sammenhæng mellem resultaterne i 8. klasses nationale test i dansk og karakterer for folkeskolens 9. klasses-prøve, end der er for 6. klasses nationale test i dansk. Dette er også forventeligt, da 8. klasse kun er ét år før 9. klasse, og dermed er der færre år og færre usikre parametre, der kan spille ind på sammenhængen mellem karaktererne i 9. klasse og resultaterne fra de nationale test i 8. klasse.

Af Figur 3.1 fremgår sammenhængen mellem samlet dansk karakter i folkeskolens 9. klasses-prøve, samlede profilområder i nationale test i dansk (læsning) i 8. klasse samt profilområder. Som det fremgår af figuren, er der tale om meget lineære sammenhænge mellem karaktererne i dansk og de nationale test i 8. klasse. Dette fremgår på alle profilområder. Figuren viser, at en elev, der eksempelvis scorede blandt de 20-29 % nederste i de nationale test i dansk (læsning), i gennemsnit opnåede en karakter i dansk på godt 6, når man ser på profilområdet "Afkodning" efter genberegningen af sværhedsgraderne i dansk (læsning). Tilsvarende fik elever i gennemsnit en karakter på godt 9, hvis de scorede blandt de 90-100 % højeste på de nationale test for samme profilområde efter genberegning af sværhedsgraderne i dansk (læsning). Figuren viser endvidere, at der er forskelle på sammenhængen mellem karakterer i folkeskolens 9. klasses-prøver i dansk og de nationale test før og efter genberegningen af de nationale tests sværhedsgrad. Forskellene er dog små, men statistisk signifikante for alle opdelinger.

Figur 3.1 Sammenhæng mellem samlet dansk karakter i folkeskolens 9. klasses-prøve, samlede profilområder i nationale test i dansk (læsning) i 8. klasse samt profilområder. Opdelt på før og efter genberegning af de nationale tests sværhedsgrad. Percentiler baseret på logit-scorer fra de nationale test.



Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

I Bilag 1 fremgår sammenhænge mellem de nationale test og karakterer fra folkeskolens 9. klasses-prøve uden opdelinger i beregninger før og efter genberegninger af sværhedsgrader. Dette er for alle nationale test, der findes opdelinger for. Som det fremgår af disse tabeller, så ligger de på nogenlunde samme niveau som ved opdelingerne, hvilket indikerer, at ændringerne i de nationale tests sværhedsgrader ikke har substantiel betydning.

4 Matematik

I det følgende vises resultaterne af sammenhængen mellem henholdsvis matematik på 3. klassetrin (Tabel 4.1), 6. klassetrin (Tabel 4.2) og de samlede karakterer i matematik i 9. klasse, dvs. gennemsnittet af karakterer fra matematikprøver med og uden hjælpemidler og karakterer fra matematikprøver med og uden hjælpemidler.

For 3. klasse er der ikke nok observationer til, at det er muligt at skelne mellem før og efter genberegningen af de nationale test i matematik i 3. klasse. Det er muligt for 6. klassetrin, og resultater for disse sammenhænge vises derfor med sammenhængene før og efter genberegningen af de nationale test i 2015.

4.1 Matematik i 3. klasse

Tabel 4.1 Korrelationer mellem de nationale test i matematik ved 3. klassetrin samt profilområder og karakterer i folkeskolens 9. klasses-prøver i matematik.

Karakterer	Matematik	Tal og algebra	Geometri	Matematik i anvendelse
Samlet matematik i 9. klasse	0,618***	0,536***	0,526***	0,561***
Prøve uden hjælpemidler	0,611***	0,534***	0,513***	0,558***
Prøve med hjælpemidler	0,563***	0,486***	0,486***	0,508***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 149.850.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

4.2 Matematik i 6. klasse

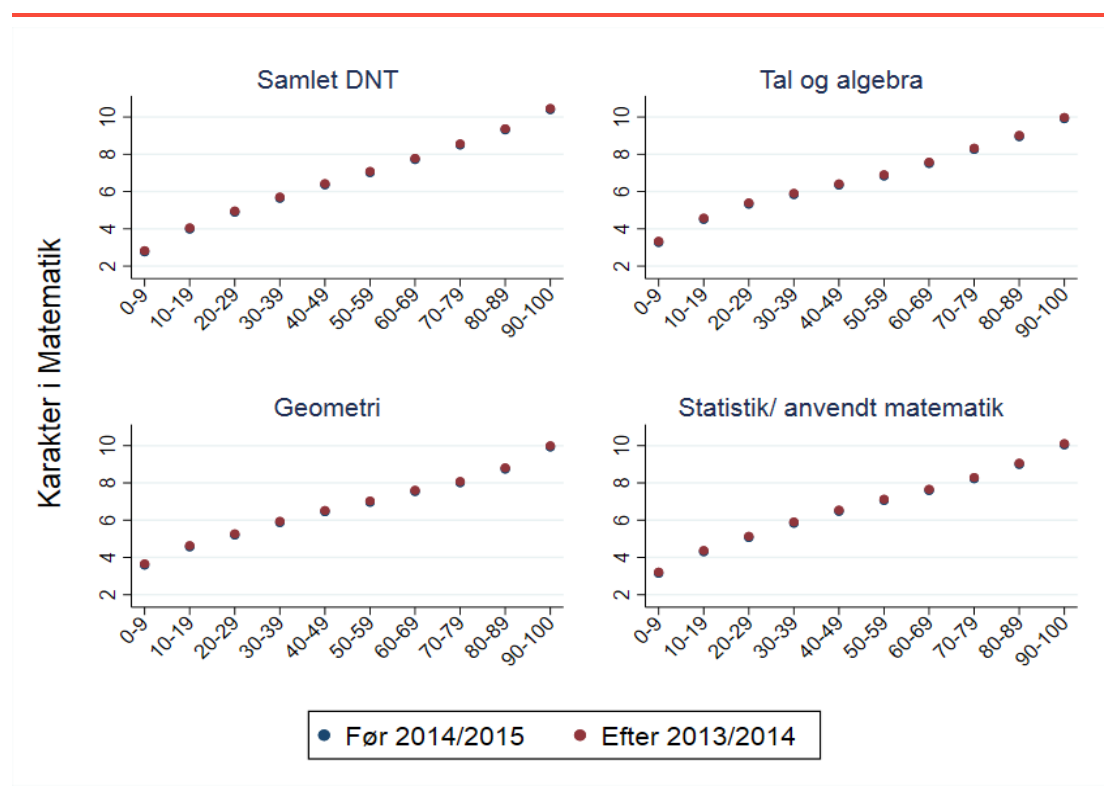
Tabel 4.2 Korrelationer mellem de nationale test i matematik ved 6. klassetrin, samt profilområder, og karakterer i folkeskolens 9. klasses-prøver i matematik, samt standpunktskarakterer i 8. klasse. Før og efter genberegning af sværhedsgrader.

Karakterer	Nationale test før skoleåret 2014/2015				Nationale test efter skoleåret 2013/2014			
	Matematik	Tal og algebra	Geometri	Matematik i anvendelse	Matematik	Tal og algebra	Geometri	Matematik i anvendelse/Statistik og sandsynlighed
Samlet matematik i 9. klasse	0,714***	0,602***	0,575***	0,635***	0,735***	0,636***	0,607***	0,654***
Prøve uden hjælpemidler	0,698***	0,593***	0,556***	0,622***	0,724***	0,629***	0,592***	0,648***
Prøve med hjælpemidler	0,658***	0,551***	0,535***	0,583***	0,674***	0,580***	0,563***	0,596***
Samlet matematik i 8. klasse	-	-	-	-	0,715***	0,629***	0,613***	0,637***
Standpunktskarakter uden hjælpemidler i 8. klasse	-	-	-	-	0,710***	0,626***	0,606***	0,635***
Standpunktskarakter med hjælpemidler i 8. klasse	-	-	-	-	0,666***	0,585***	0,574***	0,592***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge i mellem de nationale test og karakterer.
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 203.997 før skoleåret 2014/2015 og 100.488 efter skoleåret 2013/2014. For sammenhænge med standpunktskarakterer i 8. klasse er antallet af observationer 52.036. Der er statistisk signifikante forskelle på alle profilområder og de samlede mål for nationale test i matematik før og efter genberegning af fagets sværhedsgrader.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Figur 4.1 Sammenhæng mellem samlet matematikkarakter i folkeskolens 9. klassesprøve, samlede profilområder i nationale test i matematik i 6. klasse samt profilområder. Opdelt på før og efter genberegning af de nationale tests sværhedsgrad. Percentiler baseret på logit-scoringer fra de nationale test.



Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

5 Biologi i 8. klasse

Biologi er et udtræksfag i 9. klasse, og der er derfor ikke lige så mange elever, som både har taget de nationale test i biologi og karakterer ved folkeskolens 9. klasses-prøver, som der er ved øvrige fag. Resultaterne for biologi i 8. klasses nationale test er opdelt i perioden før og efter genberegning af testen og profilområdernes sværhedsgrad i 2015 – og fremgår af Tabel 5.1.

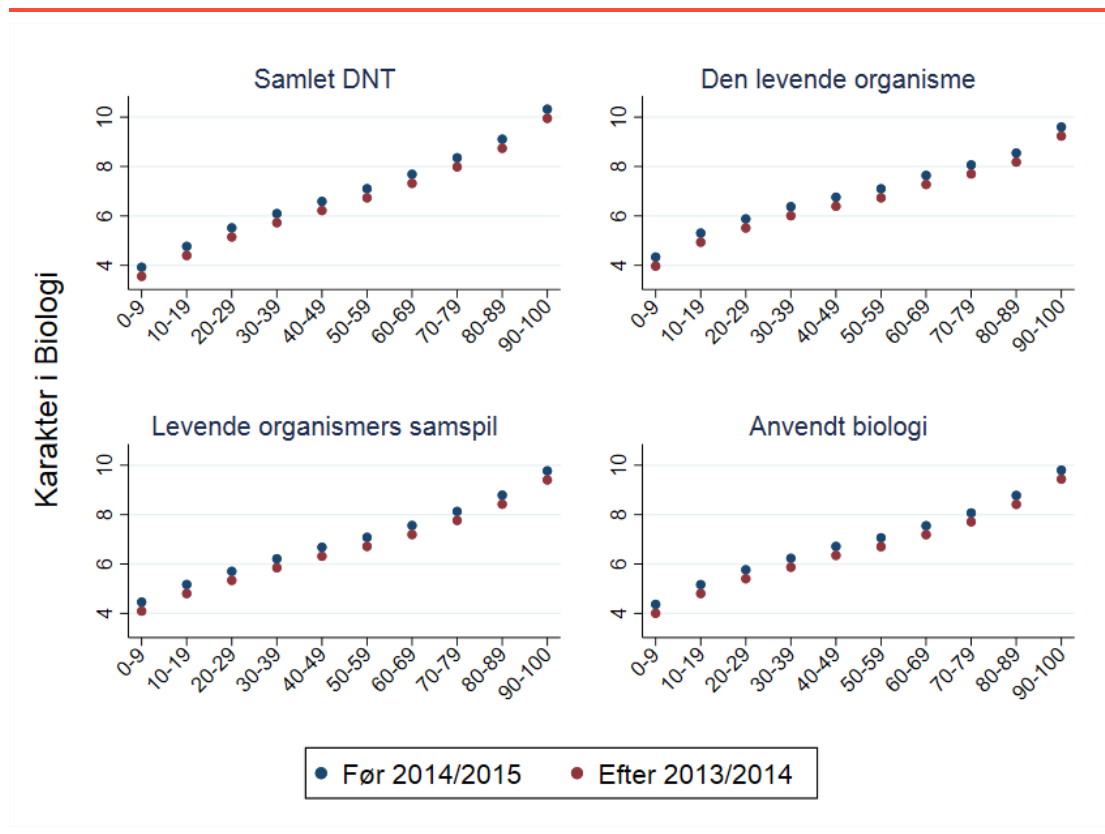
Tabel 5.1 Korrelationer mellem de nationale test i biologi ved 8. klassetrin og karakterer i folkeskolens 9. klasses-prøver i biologi. Før og efter genberegning.

Karakterer	Nationale test før skoleåret 2014/2015				Nationale test efter skoleåret 2013/2014			
	Biologi	Den levende organisme	Levende organismers samspil	Anvendt biologi	Biologi	Den levende organisme	Levende organismers samspil	Anvendt biologi
Biologi i 9. klasse	0,632***	0,503***	0,534***	0,531***	0,660***	0,518***	0,544***	0,553***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer. * p < 0.05, ** p < 0.01, *** p < 0.001. Antallet af observationer er 77.427 før skoleåret 2014/2015 og 43.142 efter skoleåret 2013/2014. Der er statistisk signifikante forskelle på alle profilområder og de samlede mål for nationale test i biologi før og efter genberegning af fagets sværhedsgrader.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Figur 5.1 Sammenhæng mellem karakter i folkeskolens 9. klasses-prøve i biologi, samlede profilområder i nationale test i biologi i 8. klasse samt profilområder. Opdelt på før og efter genberegning af de nationale tests sværhedsgrad. Percentiler baseret på logit-scorer fra de nationale test.



Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

6 Fysik/kemi i 8. klasse

Karaktererne i folkeskolens 9. klasses-prøve er baseret på de bundne prøver i fysik/kemi i 9. klasse frem til skoleåret 2016/2017. Udtræksprøverne i fysik/kemi er ikke medtaget, da disse vil kunne skævvride resultaterne, da nogle elever vil have taget mere end én afgangsprøve.

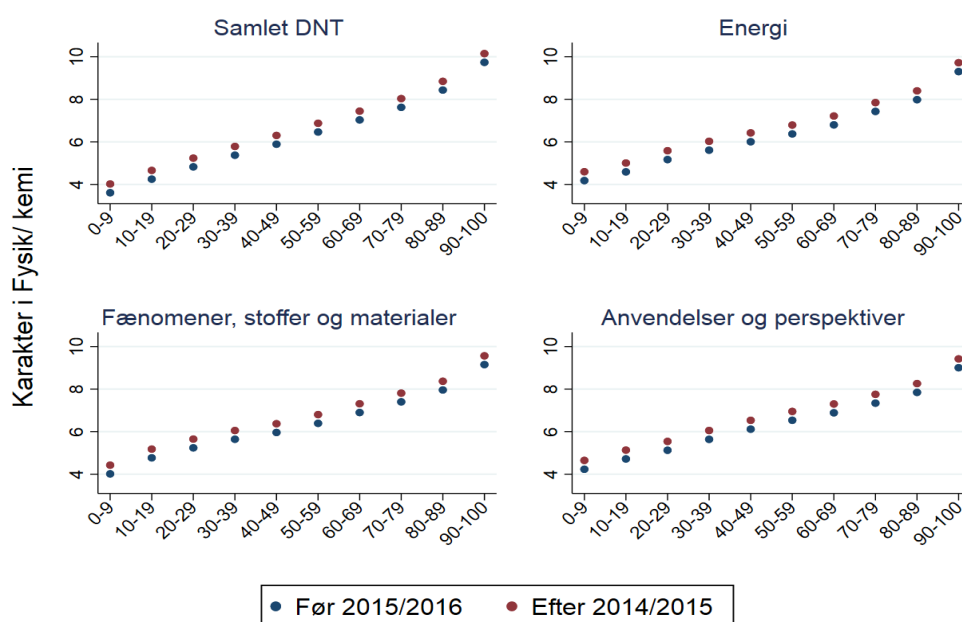
Tabel 6.1 Korrelationer mellem de nationale test i fysik/kemi ved 8. klassetrin, samt profilområder, og karakterer i folkeskolens 9. klasses-prøver i fysik/kemi. Før og efter genberegning af sværhedsgrader.

Karakterer	Nationale test før skoleåret 2015/2016				Nationale test efter skoleåret 2014/2015			
	Fysik/kemi	Energi	Fænomener, stoffer og materialer	Anvendelser og perspektiver	Fysik/kemi	Energi	Fænomener, stoffer og materialer	Anvendelser og perspektiver
Fysik/kemi i 9. klasse	0,494***	0,406***	0,396***	0,381***	0,474***	0,403***	0,391***	0,375***
Fysik/kemi i 8. klasse					0,602***	0,504***	0,485***	0,462***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer. * p < 0.05, ** p < 0.01, *** p < 0.001. Antallet af observationer er 231.027 før skoleåret 2015/2016 og 184.329 efter skoleåret 2014/2015. Antallet af observationer for 8. klasses standpunktskarakterer er 47.570. Der er ikke statistisk signifikante forskelle for profilområder og de samlede mål for nationale test i biologi og sammenhængen med karakterer i folkeskolens 9. klasses-prøver før og efter genberegning af fagets sværhedsgrader.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Figur 6.1 Sammenhæng mellem karakterer i folkeskolens 9. klasses-prøve i fysik/kemi, samlede profilområder i nationale test i fysik/kemi i 8. klasse samt profilområder. Opdelt på før og efter genberegning af de nationale tests sværhedsgrad. Percentiler baseret på logit-scorer fra de nationale test.



Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

7 Geografi 8. i klasse

Geografi er et udtræksfag i folkeskolens 9. klasses-prøver. Der er derfor ikke lige så mange elever, som både har taget de nationale test i geografi og aflagt prøver ved folkeskolens 9. klasses-prøver, som der er ved øvrige fag, hvilket – i lighed med de øvrige udtræksfag – kan give en usikkerhed i forhold til resultaterne for geografi. Resultaterne for geografi i 8. klasses nationale test er opdelt i perioden før og efter genberegning af testen i 2015.

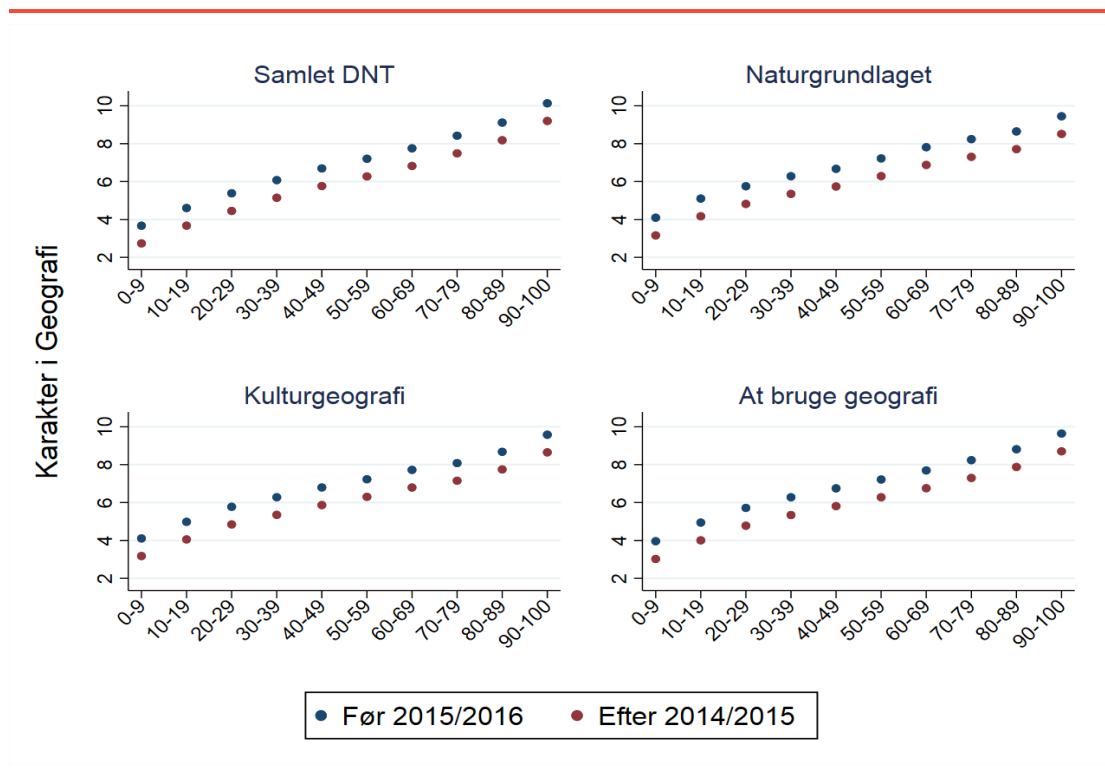
Tabel 7.1 Korrelationer mellem de nationale test i geografi ved 8. klasses-trin samt profilområder og karakterer i folkeskolens 9. klasses-prøver i geografi. Før og efter genberegning af sværhedsgrader.

Karakterer	Nationale test før skoleåret 2015/2016				Nationale test efter skoleåret 2014/2015			
	Geografi	Naturgrundlaget	Kultur-geografi	At bruge geografi	Geografi	Naturgrundlaget	Kultur-geografi	At bruge geografi
Geografi i 9. klasse	0,628***	0,515***	0,517***	0,546***	0,667***	0,546***	0,559***	0,566***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$. Antallet af observationer er 88.952 før skoleåret 2015/2016 og 28.761 efter skoleåret 2014/2015. Der er statistisk signifikante forskelle på alle profilområder og det samlede mål for de nationale test i geografi før og efter genberegning af fagets sværhedsgrader.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Figur 7.1 Sammenhæng mellem karakter i folkeskolens 9. klasses-prøve i geografi, samlede profilområder i nationale test i geografi i 8. klasse samt profilområder. Opdelt på før og efter genberegning af de nationale tests sværhedsgrad. Percentiler baseret på logit-scoringer fra de nationale test.



Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

8 Engelsk 7. i klasse

Det følgende viser sammenhænge mellem de nationale test i engelsk, 7. klasse og karaktererne ved folkeskolens 9. klasses-prøve i mundtlig engelsk. Karaktererne i folkeskolens 9. klasses-prøve er baseret på de bundne prøver i engelsk i 9. klasse. Den skriftlige prøve i engelsk er ikke medtaget, da den vil kunne skævvride resultaterne, fordi nogle elever vil have taget mere end én afgangsprøve.

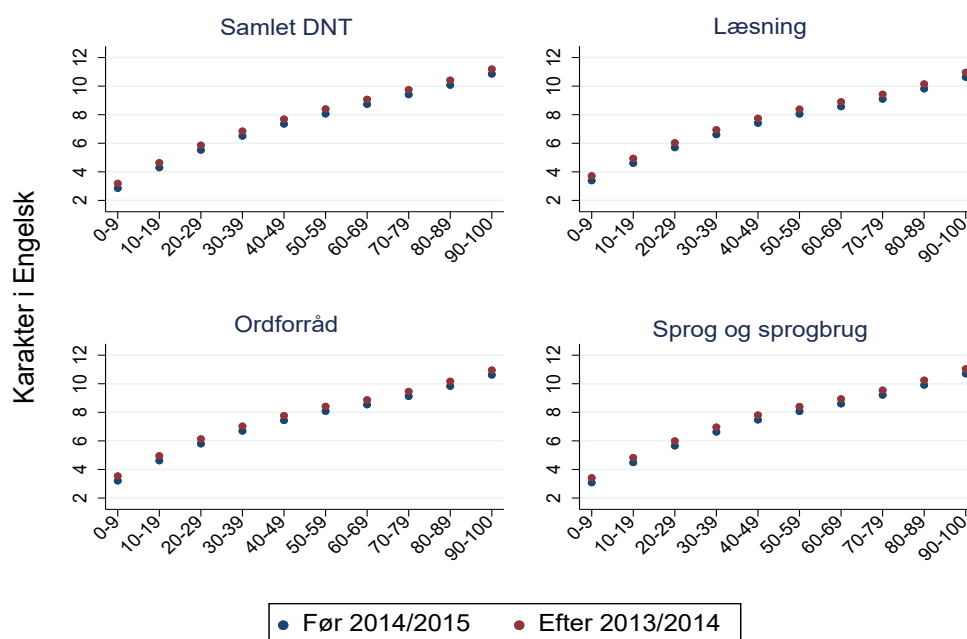
Tabel 8.1 Korrelationer mellem de nationale test i engelsk ved 7. klassetrin og karakterer i folkeskolens 9. klasses-prøver i engelsk. Før og efter genberegning af sværhedsgrader.

Karakterer	Nationale test før skoleåret 2014/2015				Nationale test efter skoleåret 2013/2014			
	Engelsk	Læsning	Ordforråd	Sprog og sprogbrug	Engelsk	Læsning	Ordforråd	Sprog og sprogbrug
Engelsk i 9. klasse	0,666***	0,594***	0,601***	0,622***	0,670***	0,610***	0,611***	0,636***
Engelsk i 8. klasse	-	-	-	-	0,733***	0,680***	0,664***	0,696***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 190.874 før skoleåret 2014/2015 og 142.631 efter skoleåret 2013/2014. Antallet af observationer for standpunktskarakterer i 8. klasse er 50.012. Der er ikke statistisk signifikante forskelle mellem det samlede mål for engelsk før og efter genberegning af sværhedsgrader. For profilområderne "læsning", "ordforråd" og "sprog og sprogbrug" er der statistisk signifikante forskelle.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Figur 8.1 Sammenhæng mellem karakterer i engelsk i folkeskolens 9. klasses-prøve, samlede profilområder i nationale test i engelsk samt profilområder. Opdelt på før og efter genberegning af de nationale tests sværhedsgrad. Percentiler baseret på logit-scorer fra de nationale test.



Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Litteratur

- Beuchert, L.V. & Nandrup, A.B. (2018). The Danish National Tests at a Glance. *Nationaløkonomisk Tidsskrift*, 2018(1), 1-37.
- Brookhart, S.M., Guskey, T.R., Bowers, A.J., McMillan, J.H., Smith, J.K., Smith, L.F., Welsh, M.E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86(4), 803–848.
- Bundsgaard, J. & Kreiner, S. (2019). *Undersøgelse af De Nationale Tests måleegenskaber. 2. udgave*. København: Danmarks Institut for Pædagogik og Uddannelse, Aarhus Universitet.
- Børne- og Undervisningsministeriet (2019). *Om de nationale test* <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test--trivselsmaaling-og-sprogproever/nationale-test/om-de-nationale-test>, hjemmeside tilgået 2019 10 14
- Calmar Andersen, S. & Nielsen, H.S. (2019). Learning from Performance Information. *Journal of Public Administration Research And Theory*, 2009, 1-17.
- Cousins, J.B. & Shulha, L.M. (2006). A Comparative Analysis of Evaluation Utilization and its Cognate Field of Inquiry: Current Issue and Trends. I: Shaw, I.F., Greene, J.C. & Mark, M.M. (red.) *The Sage Handbook of Evaluation*. London: SAGE.
- DAMVAD (2014). *PISA-relateret af de kriteriebaserede nationale test. Delrapport 1 – formidling af resultater*. København: DAMVAD.
- Flarup, L.H. (2020). *Evalueringen af de nationale test. Tværgående evalueringsrapport*. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.
- Greene, J.G. (1988), Stakeholder Participation and Utilization in Program Evaluation Evaluation Review, vol. 12, nr. 2, pp. 91-116.
- Jennrich, R.I. (1970). An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association*, 65(330), 904–912.
- Kristensen, Nicolai & Skov, Peter Rohde (2019): *Betydningen af kompetencedækning og læreruddannelsesbaggrund*. København: VIVE – Det Nationale Forsknings og Analysecenter for Velfærd.
- Ministeriet for Børn, Undervisning og Ligestilling (2016). *De nationale tests måleegenskaber*. København: Ministeriet for Børn, Undervisning og Ligestilling.
- Ministeriet for Børn, Undervisning og Ligestilling (2016b). *Bilag 2: Undersøgelse af de nationale tests reliabilitet*. København: Ministeriet for Børn, Undervisning og Ligestilling.
- Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Pett, M.A., Lackey, N.R. & Sullivan, J.J. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. California: SAGE.

- Preskill, H., Zuckerman, B. & Matthews, B. (2003). An Exploratory Study of Process Use: Findings and Implications for Future Research. *American Journal of Evaluation*, 24(4), 423-442.
- Preskill, H. & Boyle, S. (2008). A Multidisciplinary Model of Evaluation Capacity Building. *American Journal of Evaluation*, 29(4), 443-459.
- ROCKWOOL Fonden (2019). *Børns udvikling af sproglige færdigheder har social slagside*. Nyt fra RFF. København: ROCKWOOL Fondens Forskningsenhed.
- Sievertsen, H.H., Gino, F. & Piovesan, M. (2016). Cognitive fatigue influences students' performance on standardized tests. *Proceedings of the National Academy of Sciences*, 113(10), 2621-2624.
- Undervisningsministeriet (2005). *Lov om ændring af lov om folkeskolen L101*. København: Undervisningsministeriet.
- Undervisningsministeriet (2006). *Lov om ændring af lov om folkeskolen L170*. København: Undervisningsministeriet.
- Undervisningsministeriet (2015). *Den adaptive algoritme i De Nationale Test*. Aarhus: Undervisningsministeriet.
- Undervisningsministeriet (2018). *Vejledning om de nationale test - til skoleledere*. København: Undervisningsministeriet.
- Wiberg, M. (2019). The relationship between TIMSS mathematics achievements, grades, and national test scores. *Education Inquiry*, 10(4), 328-343.

Bilag 1 Tabeller uden opdeling

De følgende tabeller viser korrelationerne mellem de nationale test og karakterer i folkeskolens 9. classes-prøver. I forbindelse med ændringen af profilområdet "Matematik i anvendelse" til "Statistik og sandsynlighed", så fremgår disse sammenhænge i tabellerne under Matematik. Det skyldes, at denne ændring i profilområderne er indført samtidig med genberegningerne af sværhedsgraderne. Tilsvarende fremgår sammenhænge mellem standpunktskarakterer i 8. klasse og de nationale test også i hovedtabellerne, da oplysninger om standpunktskarakterer kun er tilgængelige for et enkelt skoleår (2017/2018).

Bilagstabel 1.1 Korrelationer mellem de nationale test i dansk (læsning) ved 6. klasses-trin og karakterer i folkeskolens 9. classes-prøver i dansk.

Karakterer	Nationale test			
	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse
Samlet dansk, 9. klasse	0,688***	0,627***	0,642***	0,522***
Læsning, 9. klasse	0,628***	0,587***	0,559***	0,490***
Mundtlig, 9. klasse	0,435***	0,415***	0,373***	0,344***
Retskrivning, 9. klasse	0,665***	0,574***	0,678***	0,483***
Skriftlig, 9. klasse	0,515***	0,464***	0,490***	0,386***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 481.406.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Bilagstabel 1.2 Korrelationer mellem de nationale test i dansk (læsning) ved 8. klasses-trin og karakterer i folkeskolens 9. classes-prøver i dansk.

Karakterer	Nationale test			
	Dansk (læsning)	Tekstforståelse	Afkodning	Sprogforståelse
Samlet dansk, 9. klasse	0,696***	0,667***	0,432***	0,416***
Læsning, 9. klasse	0,629***	0,618***	0,389***	0,402***
Mundtlig, 9. klasse	0,457***	0,460***	0,251***	0,274***
Retskrivning, 9. klasse	0,670***	0,602***	0,446***	0,383***
Skriftlig, 9. klasse	0,536***	0,510***	0,368***	0,337***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 449.082.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Bilagstabel 1.3 Korrelationer mellem de nationale test i matematik ved 6. klasses-trin, samt profilområder og karakterer i folkeskolens 9. classes-prøver i matematik.

Karakterer	Nationale test		
	Matematik	Tal og algebra	Geometri
Samlet matematik i 9. klasse	0,721***	0,613***	0,585***
Prøve uden hjælpemidler	0,706***	0,604***	0,567***
Prøve med hjælpemidler	0,662***	0,560***	0,543***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 481.030.

Bilagstabel 1.4 Korrelationer mellem de nationale test i biologi ved 8. klassetrin og karakterer i folkeskolens 9. klasses-prøve i biologi.

Karakterer	Nationale test			
	Biologi	Den levende organisme	Levende organismer samspil	Anvendt biologi
Biologi i 9. klasse	0,641***	0,508***	0,536***	0,539***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 338.486.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Bilagstabel 1.5 Korrelationer mellem de nationale test i fysik/kemi ved 8. klassetrin samt profilområder og karakterer i folkeskolens 9. klasses-prøve i fysik/kemi.

Karakterer	Nationale test			
	Fysik/kemi	Energi	Fænomener, stoffer og materialer	Anvendelser og perspektiver
Fysik/kemi i 9. klasse	0,493***	0,406***	0,397***	0,381***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 436.923.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Bilagstabel 1.6 Korrelationer mellem de nationale test i geografi ved 8. klassetrin samt profilområder og karakterer i folkeskolens 9. klasses-prøve i geografi.

Karakterer	Nationale test			
	Geografi	Naturgrundlaget	Kulturgeografi	At bruge geografi
Geografi i 9. klasse	0,631***	0,515***	0,523***	0,544***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 388.545.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

Bilagstabel 1.7 Korrelationer mellem de nationale test i engelsk ved 7. klassetrin og karakterer i folkeskolens 9. klasses-prøve i engelsk.

Karakterer	Nationale test			
	Engelsk	Læsning	Ordforråd	Sprog og sprogbrug
Engelsk i 9. klasse	0,668***	0,601***	0,605***	0,627***

Anm.: Signifikansstjerner henviser til statistisk signifikante sammenhænge mellem de nationale test og karakterer.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Antallet af observationer er 333.505.

Kilde: Data kommer fra STILs registre for nationale test og Danmarks Statistiks registre. Egne beregninger.

VIDEN
VELFÆRD

DET NATIONALE FORSKNINGS-
OG ANALYSECENTER FOR VELFÆRD