

Rapport

Evaluering af ordningen med én bedømmer ved folkeskolens skriftlige prøver



Lasse Hønge Flarup, Louise Ladegaard Bro og Kasper Skou Arendt

Evaluering af ordningen med én bedømmer ved folkeskolens skriftlige prøver

© VIVE og forfatterne, 2019

e-ISBN: 978-87-7119-635-1

Forsidefoto: Ricky John Molloy

Projekt: 301234

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.

Forord

I forbindelse med Lov om ændring af lov om folkeskolen (lov nr. 747 af 1. juni 2015) blev bedømmelsesordningen ved skriftlige prøver på 9. og 10. klassetrin ændret, så de fremover kun bedømmes af en enkelt person, nemlig én statsligt beskikket censor. Ordningen med én bedømmer ved skriftlige prøver har været brugt siden maj 2016.

I denne rapport undersøges det, hvordan karaktererne i de censorbedømte skriftlige (del-)prøver i 9. og 10. klasse har udviklet sig over tid og i sammenligning med standpunktskaraktererne og prøvekarakterer i samme fag men med mundtlig prøveform. Det undersøges også, hvordan bedømmelsesordningen vurderes af henholdsvis censorer, skoleledere og skolelærere, herunder hvordan de forskellige grupper vurderer muligheden for at give skolelærerne skriftligt feedback under den nye ordning. Rapporten belyser også, hvordan censorerne, skolelederne og skolelærerne vurderer censorrollen, samt hvad skolelederne gør for at finde og kvalificere censorer til bedømmelsesarbejdet. Endelig undersøges det, hvordan samarbejdet mellem censor og skole opleves under den nye ordning, samt hvordan censorerne anvender og vurderer de eksisterende bedømmelsesværktøjer, censorkurser, webinarer og rettegrupper.

Rapporten bygger på en analyse af karakterdata fra Undervisningsministeriets databank, spørgeskemaundersøgelser blandt beskikkede censorer og skoleledere samt kvalitative interview med henholdsvis beskikkede censorer, skoleledere og skolelærere.

Vi takker for værdifulde kommentarer fra vores eksterne reviewere. Vi takker endvidere de mange beskikkede censorer, skoleledere og skolelærere, som deltog i de kvalitative interview, og som besvarede spørgeskemaerne til brug for undersøgelsen. Undersøgelsen er iværksat på foranledning af Styrelsen for Undervisning og Kvalitet.

Hanne Søndergård Pedersen

*Konst. forsknings- og analysechef, VIVE Børn og uddannelse
2019*

Indhold

Sammenfatning	5
Evalueringens resultater	5
1 Indledning	9
1.1 Baggrund	9
1.2 Evalueringens formål	10
1.3 Læsevejledning	11
2 Udvikling i prøvekarakterer	12
2.1 Udvikling i karaktergivning i skriftlige (del-)prøver	12
2.2 Udvikling af karaktergivning med standpunktskarakterer som benchmark	14
2.3 Udvikling af karaktergivning med mundtlige prøvekarakterer som benchmark	17
2.4 Delkonklusion	17
3 Vurderingen af den nye bedømmelsesordning	19
3.1 Oplevelser og erfaringer med den nye bedømmelsesordning	19
3.2 Feedback	23
3.3 Delkonklusion	28
4 Censorernes arbejdsvilkår og kvalificering af deres arbejde	30
4.1 Vurderingen af censorrollen	30
4.2 Skoleledernes arbejde for at finde nye censorer og kvalificere deres arbejde	32
4.3 Samarbejdet mellem censorer og skolerne	33
4.4 Bedømmelsesværktøjer	35
4.5 Rettegrupper og sparring	38
4.6 Censorkurser og webinarer	42
4.7 Delkonklusion	45
Bilag 1 Metode	47
Bilag 2 Bilagstabeller	52
Bilag 3 Tabeller og figurer til karakteranalyse	61
Litteratur	69

Sammenfatning

VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd har for Styrelsen for Undervisning og Kvalitet (herefter STUK) evalueret ordningen med én bedømmer ved folkeskolens skriftlige prøver. Bedømmelsesordningen blev til lov i 2015 og betyder, at elever ved skriftlige prøver på 9. og 10. klassetrin kun bedømmes af en enkelt person nemlig én statsligt beskikket censor.

I forbindelse med ordningens implementering blev der nedsat en følgegruppe, der fulgte ordningen i to år. Evaluering følger op og bygger videre på følgegruppens konklusioner og anbefalinger. Formålet med evalueringen er, at undersøge, hvordan den nye ordning med én bedømmer ved de skriftlige prøver fungerer og opleves.

Konkret undersøger VIVE:

Karakterudvikling	Hvordan har karaktergivningene ved de skriftlige (del-)prøver udviklet sig før og efter indførelsen af énbedømmerordningen?
Vurdering af den nye bedømmelsesordning	Hvordan oplever censorerne, skolelederne, og skolelærerne den nye bedømmelsesordning? Hvordan vurderer censorerne, skolelederne og skolelærerne muligheden for at give skolelærerne feedback under den nye bedømmelsesordning?
Censorrollen og kvalificeringen af censorarbejdet	Hvordan vurderer censorerne, skolelederne og skolelærerne censorrollen? Hvad gør skolelederne for at finde og kvalificere censorerne til erhvervet? Hvordan opleves samarbejdet mellem censor og skole af censorerne, skolelederne og skolelærerne, herunder i forbindelse med klager, og når skolerne stiller spørgsmål ved censurs bedømmelse? Hvordan anvender censorerne de eksisterende bedømmelsesværktøjer, censorkurser og rettegrupper?

Evalueringen fokuserer endvidere på forskelle mellem skoletyper: folkeskoler, frie grundskoler og efterskoler, og på forskelle mellem fagene under ordningen: matematik, dansk, engelsk, tysk og fransk.

Evalueringens resultater

Herunder præsenteres evalueringens overordnede konklusioner og resultater. Evalueringens resultater bygger på VIVEs analyse af karakterdata fra Undervisningsministeriets databank, spørgeskemaundersøgelser blandt beskikkede censorer og skoleledere samt kvalitative interview med beskikkede censorer, skoleledere og lærere.

Marginale ændringer i karakterudviklingen efter indførelse af énbedømmerordningen

Karakterniveauet for de skriftlige prøver har ligget stabilt eller er steget en lille smule i perioden fra 2012 til 2018, hvor ordningen med én bedømmer ved de skriftlige prøver er blevet indført i sommeren 2016. Udviklingen er generelt stabil, og der er ikke klare tegn på, at den positive udvikling skyldes ordningen med én bedømmer. Det skyldes, at når man sammenligner med et relevant benchmark, standpunktskarakterer i samme fag, genfindes en lignende positiv udvikling i perioden 2012-2018. Derudover viser analysen, at de skriftlige prøvekarakterer generelt ligger lidt lavere end niveauet for standpunktskarakterer både før og efter indførelsen af den nye ordning.

Det er indikationer på, at afstanden mellem de gennemsnitlige standpunkts- og prøvekarakterer i både matematik og tysk er indsnævret efter ordningens indførelse. I fransk er afstanden steget, og prøvekaraktererne kommer dermed til at ligge endnu lavere end standpunktskaraktererne.

En analyse af udviklingen i prøvekarakterer viser for skriftligt dansk, matematik og engelsk sammenholdt med et andet relevant benchmark, mundtlige prøver i samme fag, at afstanden mellem mundtlige og skriftlige prøvekarakterer er blevet marginalt, men dog signifikant, større i dansk og matematik. Det vil sige, at de mundtlige prøvekarakterer er steget mere end de skriftlige. Der er dog tale om en begrænset udvikling i afstanden. For engelsk findes ikke signifikante forskelle. En separat analyse af udviklingen i karakterspredningen mellem standpunkts- og prøvekarakterer viser, at forskellen mellem de to karaktertyper i dansk og matematik er blevet lidt større efter ordningens indførelse.

Samlet viser analysen af udviklingen i karakterniveau, at der efter ordningens indførelse er sket en række ændringer. Ændringerne er dog marginale og kan ikke entydigt tilskrives indførelsen af ordningen med én bedømmer.

Censorerne er overvejende positive over for ordningen, mens skoleledere og lærere er det modsatte

Den overordnede oplevelse af bedømmelsesordningen varierer både mellem og inden for grupperne af censorer og skoleledere. Mens der er mellem 50 og 60 % af censorerne, der oplever, at ordningen med én bedømmer fungerer godt eller særdeles godt, er dette kun tilfældet for mellem 22-33 % af skolelederne. Inden for gruppen af censorer er matematikcensorerne generelt mere positive over for bedømmelsesordningen end danskcensorerne og sprogfagscensorerne. Det kan skyldes, at der er en større oplevet subjektivitet i karaktergivningene blandt censorerne i dansk og sprog – og dermed også en tendens til større bekymring for rettepålideligheden.

Skolelederne og lærerne er bekymrede for rettepålideligheden og de svage elever

For de skoleledere, der har været ledere både før og efter indførelsen af ordningen, er langt de fleste ledere blevet mere usikre på bedømmelsernes pålidelighed. For mange af de adspurgte skoleledere forstærkes problematikken af, at karaktererne i dag betyder mere for elevernes adgang til ungdomsuddannelserne. Oplevelsen er desuden, at det især er elever i den lave ende af karakter-skalaen, som undlader at klage – selvom det særligt er dem, der kan få problemer med at komme videre i uddannelsessystemet. Især skoleledere fra skoler med lave eller middel karakterniveauer fremhæver desuden, at det er et problem, at lærerne med den nye ordning har sværere ved at opfordre eleverne og deres forældre til at klage, da læreren ikke længere har indsigt i elevens besvarelse.

Alle ønsker at styrke måderne at give feedback på

Både censorerne, skolelederne og skolelærerne udtrykker frustration over, at skolelærerne ikke længere får feedback fra censorerne. Selvom der på censorsiden også udtrykkes fordele ved ikke

at skulle forhandle med lærerne, og selvom skolelederne også kan se fordelene ved, at lærerne ikke skal have tid til at rette, er den overordnede oplevelse, at feedbacken har stor værdi. Der er forskellige holdninger til skriftlig feedback samt forskellige forslag til feedbackformat. Dog foretrækker lærerne imidlertid også den mundtlige, faglige feedback.

Censorrollen er en fordel for både skole og censor, men ikke særlig attraktiv

Både censorer og skoleledere ser det som en fordel for de enkelte lærere og for skolerne at være/have beskikkede censorer. Hvervet ses som en god mulighed for faglig opdatering og udvikling, og censorerne ses generelt som gode sparrings- og informationskilder på skolen. Alligevel er det skoleledernes vurdering, at der ikke er ret stor interesse for hvervet blandt lærerne. På tværs af både censorer, skoleledere og skolelærere er opfattelsen, at mange lærere vælger censorhvervet fra, fordi det er meget arbejde på et tidspunkt af året, hvor man i forvejen har travlt. Der er desuden flere lærere, som ikke finder opgaven interessant. For mange skoleledere er fordelene tydelige, men det er ikke nødvendigvis ensbetydende med, at de oplever et ansvar for at motivere og kvalificere lærerne til censorhvervet.

Interaktionen mellem censorer og skolerne er begrænset og fungerer overvejende godt

Langt de fleste interaktioner om praktik eller klagesager foregår gnidningsfrit og efter relativt faste procedurer. Der efterspørges dog blandt såvel censorer som skoleledere en mere fast procedure for det praktiske samarbejde med hensyn til fremsendelse af opgaver og kontaktoplysninger mv., da dette vil lette arbejdet samt gøre det nemmere at huske fra gang til gang. I forhold til klagesager er de gældende procedurer velfungerende, om end de føles lidt langsommelige i visse tilfælde. Inden for både det praktiske samarbejde og samarbejdet vedrørende klagesager er der eksempler på dårligt samarbejde, hvor den ene part typisk oplever, at den anden part ikke håndterer opgaven tilfredsstillende. Dette kan skyldes menneskelige fejl i bedømmelser eller personbårne samarbejdsvanskeligheder.

Bedømmelsesværktøjerne bruges (efter behov) og understøtter i høj grad arbejdet

Censorerne har flere bedømmelsesværktøjer, der skal understøtte deres arbejde og være med til at sikre, at opgaverne rettes pålideligt. Censorerne fra de forskellige fag anvender de tilgængelige bedømmelsesværktøjer forskelligt og tillægger dem forskellig værdi, da der er forskel på, hvilke værktøjer de vurderer, er med til at give en ensartet bedømmelsespraksis. Analysen viser, at dansk-censorerne i særlig grad anvender vurderingsskemaet, der tager udgangspunkt i den vejledende karakterbeskrivelse, mens matematikcensorer i særlig grad anvender rettevejledningen. Blandt censorerne i sprogfagene anvendes primært vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen. Fælles er dog, at censorerne generelt vurderer, at bedømmelsesværktøjerne er brugbare, og at de bidrager til større gennemsigtighed og objektivitet i forbindelse med bedømmelserne, hvilket også lægges til grund, når de vælger et konkret værktøj.

Rettegrupperne bruges i lav grad, da censorerne søger sparring andre steder

Rettegrupperne bliver kun brugt i mindre grad af censorerne. Cirka hver femte censor har en eller flere gange gjort brug af rettegrupperne. Dette skyldes ikke ringe kvalitet i rettegrupperne men i højere grad, at censorerne bruger andre former for sparring. Censorerne søger særligt sparring blandt kolleger, som de kender i forvejen og har samarbejdet med før eller på censorkonferencen, der afholdes på SkoleKom. Det er en barriere, at censorerne ikke kender hinanden i rettegrupperne. Ligeledes udtrykker flere, at de ikke oplever et behov for at bruge rettegrupperne til sparring.

Censorkurset er uundværligt, mens webinarerne bruges efter behov

STUK gennemfører censorkurser, der har til formål at kvalificere censorerne til bedømmelsen af den kommende prøve. Analysen viser, at STUKs censorkurser bliver brugt og vurderes til at være gode til at kvalificere censorernes arbejde. Censorkurset bruges både til at holde sig opdateret på fagets udvikling og den kommende prøve samt på at få input til sin egen undervisning. Flere udtrykker et ønske om, at kurset ligger tidligere på skoleåret, så indholdet i højere grad ville kunne overføres til skolerne og lærerne og dermed i højere grad indgå i undervisningen op til prøven. Derudover viser analysen, at STUKs webinarer bliver brugt i mindre grad, men at de også vurderes til at være gode. De bruges mindre, da de ikke er obligatoriske, og da mange oplever, at de har meget travlt på det tidspunkt, hvor webinarerne gennemføres. Dertil kommer, at en del censorer ikke oplever behov for at deltage i webinarerne.

Datagrundlag

Analysen af karakterudviklingen er gennemført med populationsdata fra Undervisningsministeriets karakterregistre tilgængeligt gennem uddannelsesstatistik.dk.

Der er gennemført to spørgeskemaundersøgelser til henholdsvis censorer og skoleledere. Undersøgelsen blandt censorer er udført som en totalundersøgelse blandt alle beskikkede censorer, svarende til alle 1.155 censorer med en svarprocent på 66,9. Undersøgelsen blandt efterskoleledere er udført som en totalundersøgelse, mens en simpel tilfældig stikprøve er udtrukket for ledere på henholdsvis folkeskoler og frie og private grundskoler, i alt svarende til 990 skoleledere med en svarprocent på 45,7.

Der er gennemført interview med 86 interviewpersoner fordelt på tre respondenttyper; censorer (40 pers.), skolelærere (28 pers.) og skoleledere (18 pers.).

1 Indledning

I dette kapitel præsenteres baggrunden for evalueringen, evalueringens formål og den analytiske tilgang.

1.1 Baggrund

I forbindelse med Lov om ændring af lov om folkeskolen (lov nr. 747 af 1. juni 2015) blev bedømmelsesordningen ved skriftlige prøver på 9. og 10. klassetrin ændret, så de fremover kun bedømmes af en enkelt person, nemlig én statsligt beskikket censor. Tidligere blev de skriftlige prøver *både* bedømt af faglæreren og en censor. Ordningen med én bedømmer ved skriftlige prøver har været brugt siden maj 2016. Loven har medført ændringer i censorernes arbejde med bedømmelse af prøver og deres samarbejde med skolerne, ligesom skolerne har oplevet forandringer i forhold til den tidligere model.

Censorerne skal bedømme opgavebesvarelserne i overensstemmelse med de rammer for bedømmelsen, som Styrelsen for Undervisning og Kvalitet (STUK) har meldt ud. Opgavebesvarelserne skal have en pålidelig bedømmelse, og censorerne skal give eleverne en ensartet og retfærdig behandling. Herudover skal censorerne bidrage til at påse, at prøverne i øvrigt gennemføres i overensstemmelse med de gældende regler. Censorerne skal lave en indberetning til skolens leder, hvis de på baggrund af deres arbejde formoder, at prøveafholdelsen har været fejlfyldt, eller at der har været mangler ved den forudgående undervisning eller vejledning. Undervisningsministeriet beskikker censorer efter indstilling fra skolernes leder, men kan også beskikke andre censorer end de indstillede. Censorerne skal inden for de sidste tre år have gennemført et fuldt undervisningsforløb, der fører frem til den pågældende prøve, og de skal mindst én gang hvert tredje år deltage i de censorkurser, som ministeriet afholder. Censorerne er underlagt bestemmelserne i forvaltningsloven om inhabilitet og tavshedspligt.

Rettegrupper og vejledning til bedømmelse

I forbindelse med udvalgsbehandlingen af lovforslaget og de indkomne høringsvar blev der udtrykt bekymring for, om én censor sikrede en tilstrækkelig sikker bedømmelse sammenlignet med den tidligere ordning.

Der blev derfor indført en ordning med rettegrupper, hvor op til fem censorer bliver sat sammen, og hver censor har mulighed for faglig sparring med mindst én anden censor om en konkret besvarelse. Den anden censor har mulighed for at læse besvarelsen, når det er nødvendigt for at sikre en kvalificeret dialog. Censorerne kan derudover i gruppen løbende diskutere de generelle faglige spørgsmål, der dukker op i forbindelse med varetagelsen af deres hverv.

Tilsvarende gælder det, at censorer, der udpeges til at foretage ombedømmelse i forbindelse med klage, har mulighed for at sparre med en anden censor.

I forbindelse med indførelsen af ordningen udbyggede man også vejledningen til bedømmelse i hvert prøvefag, så den indeholder en beskrivelse af, hvilke generelle elementer en bedømmelse skal indeholde for at kunne blive bedømt i forhold til de forskellige trin på karakterskalaen.

Nedsættelse af følgegruppe

Der blev nedsat en følgegruppe, der i to år fulgte bedømmelsen ved folkeskolens afgangsprøver. Følgegruppen bestod af fagpersoner. Følgegruppen udgav i april 2018 en rapport, hvor de gennemgik udvalgte aspekter i forhold til ordningen med én bedømmer (Dolin m.fl., 2018). Følgegruppen fokuserede i deres arbejde på prøverne i matematik og dansk på 9. klassetrin og undersøgte gennem en spørgeskemaundersøgelse og fokusgrupper med rettegrupper, hvordan den nye censorordning fungerer set fra censorernes side. Via genretning af et antal prøvebesvarelser analyserede de også problematikker vedrørende censorpålidelighed og anonymisering af elevbesvarelser.

Følgegruppens undersøgelse viste, at censorerne i dansk, matematik og engelsk på 9. klassetrin tog godt imod den nye ordning. Censorerne hæftede sig bl.a. ved, at bedømmelsesprocedurerne blev mere effektive og sikre, fordi censor ikke længere skal koordinere karaktergivning med eller udsættes for pres fra klassens lærer. Omvendt fandt de også en række bekymringer særligt i forhold til, at censor ikke mere gav faglig feedback til klassens lærer, elevernes manglende retssikkerhed, fx (uundgåelige) fejl i rettearbejdet, der ikke længere opdages, fordi der kun er én bedømmer, uklarhed på skolerne om praktiske forhold, klageadgang m.m. Følgegruppen undersøgte også rettepålideligheden af censorers bedømmelser. Det vil sige 1) hvorvidt én censor i to tilfælde vurderede samme opgave ens, eller 2) hvorvidt to censorer vurderede én opgave til samme karakter (Dolin m.fl., 2018)). Følgegruppen genvurderede en række allerede bedømte dansk- og matematikopgaver, og i 40 % af tilfældene blev danskopgaverne vurderet til samme karakter, mens det var tilfældet for 72 % af matematikopgaverne. Dette stemmer overens med international forskning, der viser, at rettepålideligheden generelt er lav. Samtidig fandt de, at den væsentligste udfordring, skabt af den nye ordning, var den manglende feedback fra censor til lærer. Når lærerne ikke har dialogen med censorerne, tabes der information, der kan bidrage til en formativ evaluering af lærerens undervisning (Dolin m.fl., 2018).

Følgegruppens undersøgelse bruges som analytisk afsæt og i forbindelse med tolkning af resultaterne i nærværende evaluering.

1.2 Evalueringens formål

Formålet med evalueringen er at undersøge, hvordan den nye ordning med én bedømmer ved de skriftlige prøver fungerer og opleves. Evalueringen indeholder både en analyse af karakterudviklingen efter indførelsen af den nye bedømmelsesordning og inkluderer oplevelser fra censorer fra flere forskellige fag samt fra skoleledere og skolelærere.

Formålet med evalueringen er tosidet, i og med den skal kortlægge, hvordan censorer, lærere og skoleledere oplever ordningen med én bedømmer ved folkeskolens skriftlige prøver og danne grundlag for en stillingtagen til, hvorvidt der er behov for opfølgende tiltag. Ud over censorer i dansk og matematik inddrager evalueringen også fagene engelsk, tysk og fransk. Evalueringen forholder sig dermed til alle de censorbedømte skriftlige (del-)prøver i 9. og 10. klasse:

- 9. klasse: dansk, skriftlig fremstilling; matematik med hjælpemidler; engelsk, skriftlig fremstilling; tysk, skriftlig fremstilling; fransk, skriftlig fremstilling
- 10. klasse: dansk, del b og c; matematik; engelsk, skriftlig fremstilling; tysk, skriftlig fremstilling; fransk, skriftlig fremstilling.

Ud over at belyse oplevelsen med den nye bedømmelsesordning fra flere forskellige perspektiver,

undersøger evalueringen forskelle på oplevelsen med ordningen mellem censorer fra forskellige fag, og mellem ledere og lærere fra skoler med forskellige karakteristika.

Der fokuseres endvidere særskilt på temaet feedback med udgangspunkt i følgegruppens problematisering af den manglende tilbagemelding fra censorerne til lærerne. Evalueringen går også i dybden med censorernes arbejdsvilkår og kvalificering af deres arbejde. Her ses både på årsagerne til at (fra)vælge hvervet som censorer, samt hvad det bibringer skolerne at have beskikkede censorer. Da Undervisningsministeriet beskikker censorer efter indstilling fra skolernes ledere, ser vi på, hvad skolelederne gør for at finde og kvalificere lærere til hvervet. Endelig ser vi på, hvordan censorerne oplever og benytter sig af de eksisterende bedømmelsesværktøjer, censorkurser og rettegrupper. Evalueringen har dermed tre hovedtemaer:

Evalueringens tre hovedtemaer	
Udvikling i karakterniveau	Hvordan har karaktergivningen ved de skriftlige (del-)prøver udviklet sig før og efter indførelsen af énbedømmerordningen?
Vurdering af den nye bedømmelsesordning	Hvordan oplever censorerne, skolelederne, og skolelærerne den nye bedømmelsesordning? Hvordan vurderer censorerne, skolelederne og skolelærerne muligheden for at give skolelærerne feedback under den nye bedømmelsesordning?
Censorrollen og kvalificeringen af censorarbejdet	Hvordan vurderer censorerne og skolerne censorrollen? Hvad gør skolelederne for at finde og kvalificere censorerne til erhvervet? Hvordan opleves samarbejdet mellem censor og skole af censorerne, skolelederne og skolelærerne, herunder i forbindelse med klager, og når skolerne stiller spørgsmål ved censors bedømmelse? Hvordan anvender censorerne de eksisterende bedømmelsesværktøjer, censorkurser og rettegrupper?

Evalueringens metode til at besvare disse spørgsmål og belyse disse emner er beskrevet i Bilag 1.

1.3 Læsevejledning

Andet kapitel beskriver udviklingen i karakterer fra, før ordningen trådte i kraft til og med skoleåret 2017/2018. Tredje kapitel beskriver de overordnede erfaringer og oplevelser med den nye bedømmelsesordning set fra censorernes, skoleledernes og skolelærernes perspektiver samt mulighederne og begrænsningerne i forhold til feedback mellem skoler og censorer. Fjerde kapitel analyserer vurderingen af censorrollen, skoleledernes arbejde med at finde og kvalificere censorerne til deres arbejde samt censorernes oplevelse af de redskaber og værktøjer, de har til rådighed.

I bilag 1 præsenteres den anvendte metode mere detaljeret, i Bilag 2 findes tabellerne fra den gennemførte spørgeskemaundersøgelse blandt censorer og skoleledere og i Bilag 3 findes figurerne til analysen af udviklingen i karakterer.

2 Udvikling i prøvekarakterer

I dette kapitel undersøges, hvordan indførelsen af ordningen med én bedømmer har påvirket karakterniveauet i de censorbedømte skriftlige prøver. Det gøres gennem en analyse af, hvordan karaktergivningen ved de skriftlige (del-)prøver har udviklet sig før og efter indførelsen af énbedømmerordningen, og om der kan konstateres trends i denne udvikling. Standpunktskarakterer og mundtlige prøvekarakterer i samme fag anvendes som benchmarks for at afdække, om der kan spores lignende udviklinger i andre karakterer, som ikke er påvirket af ordningen med én bedømmer. Der analyseres på udviklingen i:

- Dansk, skriftlig fremstilling
- Matematik med hjælpemidler (9. kl.)/skriftlig matematik (10. kl.)
- Engelsk, skriftlig
- Tysk, skriftlig
- Fransk, skriftlig.

2.1 Udvikling i karaktergivning i skriftlige (del-)prøver

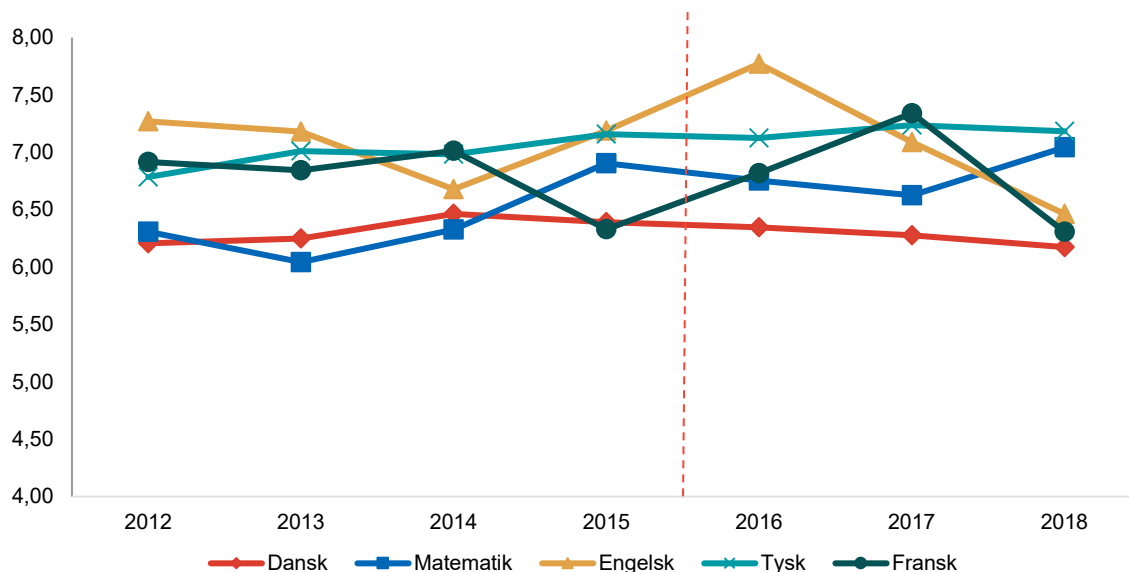
I afsnittet undersøges udviklingen i karaktergivningen til prøverne i de censorbedømte skriftlige (del-)prøver i 9. og 10. klasse¹ fra før til efter ordningens indførelse. Det undersøges også, om der er forskel på udviklingen i prøvekaraktererne på tværs af fag.

Figur 2.1 og Figur 2.2 viser den faktiske udvikling i de gennemsnitlige prøvekarakterer i ovenstående fag fra 2012 til 2018 for henholdsvis 9. og 10. klassetrin.

Udviklingen i prøvekaraktererne på 9. klassetrin fra 2012 til 2018 vises i Figur 2.1. Af figuren fremgår alle fem fag. Analysen foregår inden for det enkelte fag fra før ordningen (til og med 2015) til efter (2016 og frem). Idet prøverne i de enkelte fag er forskellige, bør man ikke sammenligne udviklingerne på tværs af fag. Af figuren fremgår det, at karakterniveauet i matematik er steget. Der er også en svag stigning i det gennemsnitlige niveau i engelsk, hvor niveauet dog daler markant efter første år med den nye ordning 2016 til 2017 og 2018. Karaktergivningen ligger i gennemsnit marginalt højere i tysk efter ordningens indførelse, mens det gennemsnitlige niveauet i dansk og fransk er stort set uændret over perioden.

¹ I sprogfag samt skriftligt dansk i 10. klasse får eleverne én samlet karakter, hvori der både indgår en digital selvrettende delopgave og en censorbedømt delopgave. I disse tilfælde har censors bedømmelse således kun indflydelse på *en del* af karakteren. Det er imidlertid ikke muligt at adskille disse i data.

Figur 2.1 Den faktiske udvikling i gennemsnitlige prøvekarakterer i skriftlige prøver i 9. klasse. Alle skoletyper. 2012-2018.

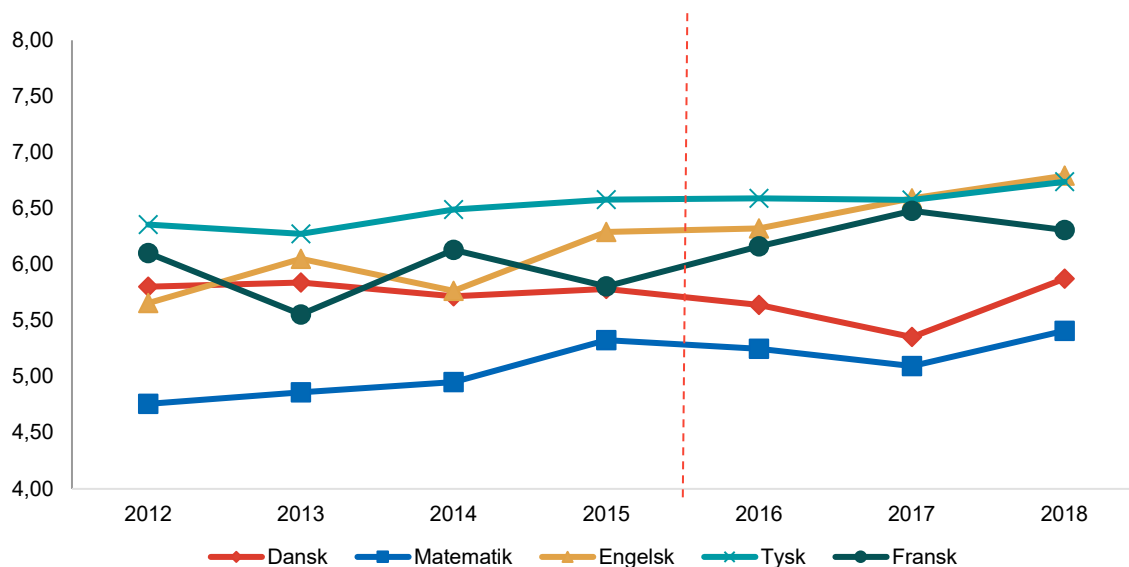


Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE.

Udviklingen i karaktergennemsnittet på 10. klassetrin er lidt mere entydigt end for 9. klassetrin. Over hele perioden stiger det gennemsnitlige niveau i matematik, fransk, engelsk og tysk. Prøvekaraktererne i dansk falder lidt i 2016 og 2017, men stiger igen i 2018 og ligger dermed på stort set samme niveau som 2012.

Figur 2.2 Den faktiske udvikling i gennemsnitlige prøvekarakterer i skriftlige prøver i 10. klasse. Alle skoletyper. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd.

Sammenligning af de to figurer viser, at karakterniveauet ved afgangsprøverne generelt ligger på et lidt højere niveau på 9. klassetrin sammenlignet med 10. klassetrin.

Figur 2.1 og Figur 2.2 viser den faktiske udvikling i karakterniveauet henover alle skoletyper, dvs. både for folkeskoler, frie og private grundskoler og efterskoler. Bilagstabel 3.1 viser en tendens til, at prøvekaraktererne stiger i alle fag på tværs af klassetrin og skoletype i perioden fra 2012-2018, med undtagelse af dansk som i gennemsnit ligger på uændret niveau. Derefter ses en tendens til, at prøvekaraktererne i dansk, matematik og tysk stiger en smule i perioden, efter ordningen blev indført, 2016-2018, mens fransk karakteren falder. For engelsk er udviklingen stort set uændret. Endelig viser bilagstabellen også, at udviklingen i prøvekaraktererne inden for dansk, matematik, fransk og tysk udvikler sig signifikant forskelligt i perioden efter indførelsen af ordningen, 2016-2018, sammenlignet med udviklingen hen over hele perioden, 2012-2018. I afsnit 2.2 undersøges ordningens betydning for karakterudviklingen nærmere.

2.2 Udvikling af karaktergivning med standpunktskarakterer som benchmark

I afsnittet undersøges udviklingen i karaktergivningen i skriftlige (del-)prøver med standpunktskaraktererne i samme fag som benchmark.

Standpunktskaraktererne anvendes alene analytisk som benchmark, og er ikke i fokus i sin egen ret. De gives umiddelbart før de skriftlige prøver og afspejler elevens faglige standpunkt på det tidspunkt. De udgør derfor et godt sammenligningsgrundlag i forhold til elevernes faglige niveau ved de skriftlige (del-)prøver, da det er en central forventning, at standpunkts- og prøvekarakterer udvikler sig parallelt. En parallel udvikling vil betyde, at ordningen ikke har haft indflydelse på prøvekaraktererne, og at udviklingen i karaktergivningen i sig selv blot er et udtryk for udviklingen i elevernes faglige præstation.

Der anvendes et difference-in-differences (DiD) design (se evt. Bilag 1), der sammenligner udviklingen i niveauet for prøvekarakterer i det enkelte fag, der er berørt af ordningen med udviklingen i standpunktskaraktererne i samme fag. Til analyserne anvendes lineære regressionsmodeller med skole-klassetrin-fixed-effects. Fixed-effects på skole- og klassetrinsniveau gør det muligt – i statistisk forstand – at tage højde for de unikke forhold, der knytter sig til det enkelte klassetrin på den enkelte skole (fx lærerens tilgang til undervisningen, classesammensætning, socioøkonomiske indikatorer, 9. eller 10. klassetrin osv.). Et sådan design sikrer, at resultaterne giver et kvalificeret bud på, hvad ordningen har betydet for den generelle udvikling i karaktergennemsnittet. Der er endvidere kontrolleret for trends i karakterudvikling forud for indførelsen af ordningen.²

I analyserne tages højde for, om karaktererne er afgivet ved de skriftlige prøver i 9. eller 10. klasse, samt om de er afgivet på folkeskoler, frie og private grundskoler eller efterskoler. De anvendte data er fra Undervisningsministeriets datavarehus og er fra skoleåret 2012/2013 til og med 2017/2018 for karaktergivning i de enkelte fag aggregeret til skoleniveau. De faktiske, gennemsnitlige karakterniveauer samt udvikling over tid for er visualiseret i Bilagsfigur 3.3-Bilagsfigur 3.7, mens DiD-analysens resultater er visualiseret i Bilagsfigur 3.8. I det følgende beskrives resultaterne af den statistiske analyse. Analysens resultater fremgår af Tabel 2.1.

² En central antagelse bag DiD-designet som model til estimering af effekter er antagelsen om "common trends". Denne handler grundlæggende om, at prøvekarakterer og standpunktskarakterer skal udvikle sig (nogenlunde) ens *forud* for ordningen. På baggrund heraf kan man beregne forskellen mellem den forventede udvikling i karakterniveau *efter* ordningen og den faktiske udvikling. DiD-analyserne, der er visualiseret i Bilagsfigur 3.8, forholder sig også til denne antagelse.

Tabel 2.1 Ændringer i karakterudvikling i karakterpoint. Skriftlige prøvekarakterer og standpunktskarakterer. Alle skoler. 2012-2018. Resultater fra lineær regression med skoleklassetrin-fixed-effects

	Dansk	Matematik	Engelsk	Fransk	Tysk
Udvikling i standpunktskarakterer over tid	0,02 ** (0,01)	0,01 * (0,01)	0,08 *** (0,01)	0,03 (0,02)	0,11 *** (0,01)
Niveauforskel: prøvekarakterer ift. standpunktskarakterer	-0,32 *** (0,02)	-0,40 *** (0,03)	-0,00 (0,03)	-0,59 *** (0,12)	0,28 *** (0,04)
Ordningens betydning for udviklingen i prøvekaraktererniveauet	0,01 * (0,01)	0,16 *** (0,01)	-0,00 (0,01)	-0,15 *** (0,05)	-0,12 *** (0,02)
Antal observationer	26.164	24.786	19.965	7.751	17.262

Anm.: * = $p < 0,10$, ** = $p < 0,05$, *** = $p < 0,01$. Standardfejl i parentes. Resultaterne er baseret på lineær regression med skoleklassetrin-fixed-effects, med kontrol for trends i udviklingen før ordningen. En falsifikationstest, hvor skæringspunktet er rykket fra 2016 til 2014 (perioden 2016 og fremefter indgår ikke) indikerer, at common trends assumption er overholdt for dansk, engelsk og tysk men ikke for matematik og fransk.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

Tabel 2.1 læses ved først at sammenligne niveauforskellen mellem prøve- og standpunktskarakterer. Dette indikerer, hvorvidt niveauet til prøverne ligger enten over eller under niveauet for standpunkt. Eksempelvis ses, at prøvekaraktererne i dansk ligger 0,32 karakterpoint under den tilsvarende standpunktskarakter. Niveauforskellen sammenlignes dernæst med estimatet for ordningens betydning: Er fortegnene for begge tal ens, bliver afstanden mellem karaktererne større over tid, mens forskellige fortegn indikerer, at karaktererne nærmer sig hinanden over tid. Her ses eksempelvis, at prøvekarakteren i dansk har nærmet sig standpunktskarakteren efter ordningens indførelse med 0,01 karakterpoint, hvilket er en meget lille ændring.

Analysen viser først, at prøvekaraktererne på tværs af de undersøgte fag generelt ligger lavere end standpunktskaraktererne, på nær i tysk. Prøvekaraktererne i dansk ligger i gennemsnit 0,32 karakterpoint under standpunktskaraktererne i dansk. For matematik og fransk ligger prøvekaraktererne i gennemsnit henholdsvis 0,40 og 0,59 point lavere end standpunktskaraktererne. I tysk ligger prøvekaraktererne i gennemsnit 0,28 point højere end standpunktskaraktererne, mens standpunkts- og prøvekarakterer i engelsk i gennemsnit ligger lige oven i hinanden, se også Bilagsfigur 3.3-Bilagsfigur 3.7

Dernæst viser analysen, at niveauet for prøvekarakterer i både matematik, fransk og tysk ændrer sig, efter ordningen med én bedømmer blev indført i forhold til benchmarket. For matematik er estimatet for prøvekarakterernes udvikling efter indførelsen af ordningen på 0,16. Dette tal har et positivt fortegn, hvilket betyder, at der er en tendens til, at prøvekarakterer bliver lidt højere, efter ordningen er blevet indført og derved nærmer sig standpunktskaraktererne, da disse ligger højere, jf. estimatet for niveauforskellen. Det samme gør sig gældende i tysk, hvor karaktererne også nærmer sig hinanden – blot med omvendt fortegn, dvs. at niveauet for prøvekaraktererne bliver lavere med et estimat på -0,12. Der er således også her en tendens til, at niveauerne for prøvekaraktererne nærmer sig benchmarket. Hvad angår fransk viser analysens resultater, at afstanden mellem standpunkts-

og prøvekarakterer bliver større, efter ordningen er blevet indført – således at standpunktskaraktererne bliver endnu højere end prøvekaraktererne med et estimat på -0,15. For dansk og engelsk ses ingen væsentlige ændringer.

2.2.1 Analyse af forskelle i prøvekaraktergivningen i forhold til benchmark

Ovenstående analyse suppleres med en undersøgelse af, om forskellen mellem prøvekarakterer og standpunktskarakterer (i absolutte tal) ændrer sig efter indførelsen af ordningen.

Mens ovenstående analyse i afsnit 2.2 giver indblik i, om indførelsen af ordningen medfører en skævvridning i de gennemsnitlige prøve- og standpunktskarakterer, så viser resultaterne ikke, hvilken karaktergivning gennemsnittet består af, og om der opstår større spredning i karaktergivningen henover karakterskalaen efter ordningens indførelse. Denne analyse undersøger netop spredningen i karaktergivningen ved at se på, om forskellen (i absolutte tal) mellem skriftlige prøvekarakterer og standpunktskarakterer ændrer sig efter indførelsen af énbedømmerordningen. Man kan have en forventning om, at de nye prøvekarakterer er mere spredte, da den nye ordning har mindre fokus på konsensus. Er dette tilfældet, kan resultaterne fortolkes således, at ordningen med én bedømmer påvirker prøvekaraktererne, men ikke om prøvekaraktererne er blevet mere eller mindre pålidelige.

Ligesom tidligere anvendes her statistiske modeller med skole-klassetrin-fixed-effects. Imidlertid foretages disse analyser alene for dansk og matematik, idet disse er bundne prøvefag. Det gør, at der findes data for hele perioden, hvilket er centralt for at sikre et solidt datagrundlag til denne type af analyse.

Tabel 2.2 Ændringer i den absolutte forskel mellem prøvekarakterer og standpunktskarakterer Skole-klassetrin-fixed-effects. 2012-2018

	Dansk	Matematik
Ændring i forskellen mellem prøvekarakter og standpunktskarakter. 2012-2018	-0,01 (0,01)	-0,03 *** (0,01)
Ændring i forskellen efter indførelsen af ordningen. 2016-2018	0,23 *** (0,02)	0,10 *** (0,02)
Antal observationer	12.888	12.093

Anm.: * = $p < 0,10$, ** = $p < 0,05$, *** = $p < 0,01$. Standardfejl i parentes. Resultaterne er baseret på lineær regression med skole-klassetrin-fixed-effects, med kontrol for trends i udviklingen før ordningen. Målet for forskellen er den absolutte forskel mellem prøvekarakterer og standpunktskarakterer. Et negativt fortegn indikerer, at forskellen bliver mindre, mens et positivt fortegn indikerer, at forskellen bliver større.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

Analysen viser, jf. Tabel 2.2, at forskellen i karaktergivningen stiger en smule i dansk og matematik i perioden, efter ordningen blev indført.³ Der er en signifikant større forskelle mellem karaktererne i dansk, 0,23 karakterpoint, og i matematik, 0,10 karakterpoint, efter ordningens indførelse.

Det betyder, at der er tendens til større forskelle i karaktergivningen mellem de to typer af karakterer efter ordningens indførelse. Analysen indikerer derfor, at ordningen har påvirket prøvekaraktergivningen. Der er dog tale om en relativt lille ændring.

³ Alternativt viser Bilagsfigur 3.9 og Bilagsfigur 3.10, hvordan forskellen mellem de to prøveformer fordeler sig henholdsvis, før og efter ordningen blev indført. Grafen for dansk viser, at der, efter ordningen er indført, oftere er små forskelle mellem karaktererne. Før ordningen var der flere større forskelle, dog sjældnere. For matematik er det grafiske billede mindre tydeligt. Fordelingen af forskelle minder her mere om hinanden end tidligere. Dog er der efter ordningen flere enkelte tilfælde med større forskelle end før ordningen.

2.3 Udvikling af karaktergivning med mundtlige prøvekarakterer som benchmark

I afsnittet undersøges udviklingen i karaktergivningen i skriftlige (del-)prøver med karaktergivningen i de mundtlige prøver i samme fag som benchmark. Dette gøres for at analysere, om der kan spores lignende udvikling den mundtlige prøveform, som ikke er påvirket af ordningen med én bedømmer.

Dette giver dels et indblik i lærerens indflydelse på karaktergivningen, dels et godt udgangspunkt for at sammenligne betydningen af ordningen for skriftlige prøver i forhold til prøveformer, der *ikke* er berørt af ordningen. Så på samme vis som i kapitlets andre analyser anvendes et Difference-in-Differences design til at undersøge, om karaktererne udvikler sig forskelligt hen over perioden. Denne analyse udføres udelukkende på data for:

- Dansk, skriftlig fremstilling og mundtlig
- Matematik med hjælpemidler (9. kl.)/skriftlig matematik (10. kl.) og mundtlig
- Engelsk, skriftlig og mundtlig.

Tabel 2.3 Ændringer i karakterudvikling i karakterpoint. Mundtlige og skriftlige prøvekarakterer. Alle skoler. 2012-2018. Resultater fra lineær regression med skole-klasse-trin-fixed-effects

	Dansk	Matematik	Engelsk
Udvikling i mundtlige prøvekarakterer over tid	0,02 ** (0,01)	0,18 *** (0,01)	0,05 *** (0,02)
Niveauforskel: skriftlige prøvekarakterer i forhold til mundtlige prøvekarakterer	-1,45 *** (0,03)	-1,47 *** (0,04)	-0,79 *** (0,07)
Ordningens betydning for udviklingen i det skriftlige prøvekarakter-niveau	-0,06 *** (0,01)	0,06 *** (0,01)	-0,04 * (0,02)
Antal observationer	19.907	19.483	9.298

Anm.: * = $p < 0,10$, ** = $p < 0,05$, *** = $p < 0,01$. Standardfejl i parentes. Resultaterne er baseret på lineær regression med skole-klasse-trin-fixed-effects.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE.

Indledende viser resultaterne i Tabel 2.3, at karaktererne i de skriftlige prøver i gennemsnit ligger på et lavere niveau end benchmarket, dvs. de mundtlige prøver. Dette fremgår af estimatet for niveauforskellen. Kombineret med estimatet for ordningens betydning viser analysen en ganske svag tendens til, at gennemsnittet i de skriftlige prøvekarakterer i dansk falder med -0,06 karakterpoint sammenlignet med de mundtlige prøvekarakterer. Det betyder med andre ord, at der er blevet større afstand mellem mundtlige og skriftlige prøvekarakterer i dansk, efter ordningen blev indført. Der er svage tegn på samme tendens for engelsk, dog er estimatet *ikke* statistisk signifikant. I matematik er der imidlertid en tendens til, at der gives lidt højere karakterer (estimat 0,06) i de skriftlige prøver sammenlignet med tidligere, hvilket indikerer, at de to karakterer nærmer sig hinanden.

2.4 Delkonklusion

På baggrund af analyserne i dette kapitel kan det overordnet konkluderes, at indførelsen af ordningen ikke tydeligt har påvirket karaktergivningen i én bestemt retning. Der er tegn på, at karakterniveauet i gennemsnit efter ordningen ligger på et lidt højere niveau end før ordningen, og at der i

enkelte fag er tegn på, at ordningen har betydet, at de gennemsnitlige standpunkts- og prøvekarakterer nærmer sig hinanden. Der er dog også tegn på, at ordningen har medvirket til at skabe større forskelle mellem lærere og censorers karaktergivninger.

Analysen af udviklingen i prøvekarakterniveauet viser for de skriftlige prøver i 9. klasse, at prøvekaraktererne er generelt steget en smule for matematik, engelsk og tysk, mens niveauet for dansk og fransk er stort set uændret. I 10. klasse ses, at karakterniveauet for de skriftlige prøver stiger i matematik, fransk, engelsk og tysk. Udviklingen i prøvekaraktererne i dansk falder lidt i 2016 og 2017, men stiger igen i 2018, og ligger dermed på stort set samme niveau som 2012. Når man alene ser på udviklingen i prøvekarakterer, ses ikke tydelige ændringer i udviklingen i forbindelse med indførelse af ordningen med én bedømmer. Henover perioden ses naturlige udsving i de enkelte år. Dette er i tråd med tidligere forskning, bl.a. Danmarks Evalueringsinstituts evaluering af 7-trinsskalaen, der viste lignende små fluktuationer i karakterniveauet fra 2007-2016 (Danmarks Evalueringsinstitut, 2019).

Når man analyserer prøvekaraktererne ud fra et benchmark, her standpunktskarakterer, ses at karakterniveauet for standpunktskarakterer overordnet ligger lidt højere end niveauet for prøvekarakterer. Og at det ser ud til, at ordningen med én bedømmer er med til at indsnævre afstanden mellem de *gennemsnitlige* standpunkts- og prøvekarakterer i både matematik og tysk, mens afstanden er stort set uændret for dansk og engelsk. For fransk viser analysen, at afstanden mellem de to *gennemsnitlige* karakterniveauer stiger, og at prøvekaraktererne dermed kommer til at ligge endnu lavere end standpunktskaraktererne efter ordningens indførelse. Supplerende analyser viser endvidere, at den nye ordning har medført flere tilfælde, hvor der er forskel på prøve- og standpunktskarakter i dansk og matematik.

Analysen af skriftlige prøvekarakterer med mundtlige prøvekarakterer som benchmark viser, at afstanden mellem mundtlige og skriftlige prøvekarakterer er blevet marginalt større i dansk og marginalt mindre i matematik. Det vil sige, at de mundtlige prøvekarakterer er steget mere end de skriftlige. Der er dog tale om en begrænset udvikling i afstanden mellem karaktererne. For engelsk findes ingen signifikante forskelle.

3 Vurderingen af den nye bedømmelsesordning

I dette kapitel undersøges, hvordan censorerne, skolelederne og skolelærerne oplever den nye bedømmelsesordning samt deres erfaringer med feedback i den nye ordning. Det undersøges ligeledes, om der er forskelle mellem censorer i forskellige fag, samt mellem skoleledere og skolelærere fra skoler med forskellige karakteristika. Konkret undersøges spørgsmålene:

- Hvordan oplever censorerne, skolelederne og skolelærerne den nye bedømmelsesordning?
- Hvordan vurderer censorerne, skolelederne og skolelærerne muligheden for at give skolelærerne feedback under den nye bedømmelsesordning?

3.1 Oplevelser og erfaringer med den nye bedømmelsesordning

I dette afsnit afdækkes censorerne, skolelederne, og skolelærernes oplevelser med den nye bedømmelsesordning. Censorerne og skoleledernes oplevelser er afdækket via en kombination af indsigter fra spørgeskemaundersøgelser og interview, mens skolelærernes oplevelser udelukkende er afdækket via interview. Afsnittet præsenterer først censorernes oplevelse af bedømmelsesordningen og derefter skoleledernes og skolelærernes oplevelse særligt med hensyn til ordningens betydning for rettepåidelighed og klager.

3.1.1 Censorernes oplevelser og erfaringer

Først vises censorernes overordnede vurdering af den nye bedømmelsesordning. Tabel 3.1 viser, at mellem 50 og 60 % af censorerne oplever, at ordningen med én bedømmer fungerer godt eller særdeles godt. Mellem 23 og 28 % mener, at ordningen fungerer nogenlunde, mens mellem 15 og 21 % af censorerne er overvejende kritiske over for ordningen. Censorernes oplevelse af ordningen fremgår samlet af Tabel 3.1.

Tabel 3.1 Censorbesvarelser på spørgsmålet: "Hvordan synes du bedømmelsesordningen (med én bedømmer) fungerer alt i alt?" Fordelt på fagtype. Procent.

	Dansk	Matematik	Sprogfag
Særdeles godt	18,1	17,1	18,3
Godt	31,9	43,3	36,6
Nogenlunde	28,4	23,2	25,8
Ikke så godt	11,4	5,7	11,8
Dårligt	6,4	7,1	5,4
Meget dårligt	3,9	3,7	2,1
Antal besvarelser	282	298	186

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,103. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne.
Kilde: VIVEs spørgeskema til censorer.

Som beskrevet i afsnit 1.1 konkluderede følgegruppen til den nye bedømmerordning, at rettepåideligheden ved én bedømmer er større i matematik end i dansk. I tabel 3.1 ses det, at matematikcensorerne i indeværende undersøgelse også er gennemsnitligt mere positive over for énbedømmerordningen end danskcensorerne og sprogcensorerne. Indsigterne fra de kvalitative interview peger

på, at der også er en større oplevet subjektivitet i karaktergivningene i dansk og sprog end i matematik. Denne oplevede subjektivitet kan give anledning til bekymring for rettespøreligheden. To censorer fortæller:

I dansk er der et mere subjektivt felt. Jeg har en frygt for – og en bekymring for – om eleverne nu kun får 'den lave karakter' [den som censorerne tidligere repræsenterede, red.], idet dialogen er afskaffet. Der er ingen forhandling længere. Jeg er bevidst om, at der er særlige 'landemærker' i dansk, som man skal pejle efter under bedømmelsesarbejdet, men disse er også subjektive. (Censor i dansk)

Eleven skal have en korrekt bedømmelse. Der er subjektivitet i faget; hvor meget vægtes indhold og grammatik? Der er Ingen gylden formular, altså er det subjektivt. (Censor i engelsk)

Selvom det især er danskens censorer og sprogcensorer, der udtrykker en oplevet subjektivitet i karaktergivningene, er der også variation inden for hvert af fagene. For eksempel er der både matematikcensorer, som mener, at bedømmelsesarbejdet er meget objektivt, mens andre mener, at matematik er et langt mere åbent og subjektivt fag i dag, end det var tidligere. Oplevelsen af bedømmelsesarbejdet kan også være betinget af, hvilket klassetrin man er censor for. For eksempel er der flere tysklærere, der fortæller, at der er større usikkerhed omkring rettespøreligheden for 10. klassernes prøver end for 9. klassernes prøver. Det skyldes, at den frie skriftlige del af prøven betyder mere for karakteren for 10. klasserne.

Der er også problemstillinger, der går igen på tværs af flere fag. Problematikken omkring rettespøreligheden gælder nemlig ikke kun i forhold til den oplevede subjektivitet i karaktergivningene; den relaterer sig også til muligheden for, at den enkelte censor laver fejl i sammentællingen af point eller i nedskrivningen af karakteren til en elev. Denne problematik nævnes fx både af censorer i matematik og tysk. Desuden er der også censorer i både matematik og tysk, som nævner udfordringen med at opdage snyd, når faglærerne ikke kontrollerer karaktererne. Som faglærer har man kendskab til, hvordan eleverne generelt præsterer, og derfor er det sværere at opdage snyd under den nye bedømmelsesordning, hvor det kun er censorerne, der ser prøverne.

3.1.2 Skoleledernes oplevelser og erfaringer

Skoleledernes oplevelser af, hvordan ordningen alt i alt fungerer, er mere negativ end censorernes, hvilket fremgår af Tabel 3.2. Her svarer mellem 22 og 33 % af skolelederne, at de synes særdeles godt eller godt om ordningen, mens mellem 30 og 44 % ikke synes så godt om ordningen – eller endda synes (meget) dårligt om ordningen. Der er ikke markante forskelle på tværs af skoletyper. Mest positive er efterskolelederne, hvor knap hver tredje skoleleder synes godt eller særdeles godt om ordningen, og hvor knap hver tredje leder er kritisk over for ordningen.

Tabel 3.2 Skolelederbesvarelser på spørgsmålet: "Hvordan synes du, at bedømmelsesordningen med én bedømmer ved folkeskolens skriftlige prøver fungerer alt i alt?" Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler
Særdeles godt	4,3	1,6	3,1
Godt	21,6	20,5	29,6
Nogenlunde	32,3	34,7	37,8
Ikke så godt	18,1	21,3	11,2
Dårligt	15,5	11,8	12,2
Meget dårligt	8,2	10,2	6,1
Antal besvarelser	232	127	98

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,409. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne.
Kilde: VIVEs spørgeskema til skoleledere.

Der kan heller ikke spores markante forskelle i interviewene med skolelederne på tværs af skoletyper. Selvom nogle af skolelederne godt kan se fordelene i, at man sparer rettetimerne for lærerne, så nævnes det på tværs af alle skoletyper, at 1) den nye bedømmelsesordning giver større usikkerhed om karaktererne, at 2) det er et problem, at lærerne ikke længere har dialogen med censorerne, og 3) at det ikke er optimalt, at det i højere grad udelukkende er elevernes og forældrenes ansvar at klage.

Fire skoleledere er specifikt udvalgt til interview, fordi de i spørgeskemaundersøgelsen har angivet, at de enten er meget tilfredse eller meget utilfredse med ordningen. Ser man på udtalelserne fra de meget utilfredse ledere, er det især (u)sikkerheden omkring bedømmelserne, der betyder noget:

Følelsesmæssigt så er oplevelsen, at det ikke er trygt, at der kun er én bedømmer. [...] Lærerne er ikke trygge, eleverne er ikke trygge, forældrene er ikke trygge. Det har været en god ting, at man vidste, at der var flere øjne på bedømmelsen. [...] (Skoleleder, meget utilfreds med ordningen)

Det er sikkerheden omkring karaktererne, der er problematisk. (Skoleleder, meget utilfreds med ordningen)

For de ledere, der er udvalgt, fordi de er meget tilfredse med ordningen, er der fokus på, at censorerne er kvalificerede til deres arbejde. Hos den ene af lederne er det netop opfattelsen, at rettevejledningerne fra ministeriet gør censorerne dygtigere.

For at undersøge skoleledernes oplevelse af pålideligheden i bedømmelserne mere dybdegående, har vi i spørgeskemaundersøgelsen bedt skolelederne om at vurdere dette emne (Tabel 3.3). Mellem 21 og 33 % af skolelederne vurderer, at ordningen fungerer godt eller særdeles godt i forhold til pålidelighed, mens mellem 35 og 46 % mener, at ordningen ikke fungerer så godt – eller endda (meget) dårligt. Mellem 33 og 39 % af skolelederne mener, at ordningen fungerer nogenlunde i forhold til rettepålidelighed. Efterskolelederne vurderer rettepålideligheden lidt mere positivt end folkeskole- og friskolelederne.

Tabel 3.3 Skolelederbesvarelser på spørgsmålet: "Hvordan vurderer du, at bedømmelsesordningen ved folkeskolens skriftlige prøver i folkeskolen fungerer i forhold til pålidelighed i bedømmelsen af elevbesvarelser (at eleverne får den rette karakter)?" Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler
Særdeles godt	2,6	2,4	4,1
Godt	26,8	18,1	28,6
Nogenlunde	39,4	33,9	32,7
Ikke så godt	15,7	25,2	18,4
Dårligt	16,2	9,5	10,2
Meget dårligt	9,4	11,0	6,1
Antal besvarelser	235	127	98

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,409. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til skoleledere.

Ser vi udelukkende på besvarelser fra de skoleledere, der har oplevet bedømmelsesarbejdet, *før* ændringen blev lavet i 2016, føler langt de fleste skoleledere, at bedømmelserne er blevet mere usikre (jf. Tabel 3.4). Faktisk svarer mellem 78 og 93 %, at de føler sig mere usikre på, at bedømmelserne af de skriftlige prøver er pålidelige

Tabel 3.4 Skolelederbesvarelser på spørgsmålet: "Når jeg sammenligner med situationen før indførelsen af ordningen med én bedømmer ved folkeskolens skriftlige prøver, føler jeg mig nu..." Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler
Mere sikker på, at bedømmelsen af de skriftlige prøver er pålidelig	13,4	4,3	14,10
Mere usikker på, at bedømmelsen af de skriftlige prøver er pålidelig	80,5	92,6	78,2
Ved ikke/kan ikke vurdere	6,2	3,2	7,7
Antal besvarelser	179	94	78

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,075. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne
 Kategorien "Mere sikker" er lagt sammen af svarkategorierne "Meget mere sikker", "Mere sikker" og "Lidt mere sikker", mens "Mere usikker" er lagt sammen af svarkategorierne "Meget mere usikker", "Mere usikker" og "Lidt mere usikker". Spørgsmålet er kun stillet til skoleledere, der har været ledere *før* den nye ordning.

Kilde: VIVEs spørgeskema til skoleledere.

Samlet set peger indsigterne fra spørgeskemaundersøgelsen og interviewene på, at det især er rettetpålideligheden, som skolelederne bekymrer sig om under den nye bedømmelsesordning. I de kvalitative interview forstærkes problematikken for flere skoleledere ved, at karakterne samtidig får større betydning for elevernes mulighed for at komme ind på de efterfølgende ungdomsuddannelser. Især for skoleledere med lave og middel karakterniveauer er det også et problem, at det i højere grad udelukkende er elevernes og forældrenes opgave at klage over karaktererne. Selvom klageproceduren ikke har ændret sig, skal lærerne ikke længere læse opgaverne, og derfor er det sværere for dem at rådgive om klager. Oplevelsen er nemlig, at eleverne i den lave ende klager mindre – selvom det specielt er dem, der kan få problemer med at komme videre i uddannelsessystemet:

Problemet er, at ordningen falder sammen med, at karakter pludselig har fået stor betydning for dem med lave karakterer [...]. Eleverne i den lave ende klager mindre, fordi de har mindre overskud og er mere ligeglade måske. Jeg synes, at systemet er okay, vi

lever med småskavankerne, men vi er nødt til at øve os i at klage, være tydelige og aktive i klagevejledningen til eleverne for at undgå, at en elev ikke kan komme ind på en ungdomsuddannelse pga. en forkert karakter. (Skoleleder på skole med middel karaktergrundlag)

3.1.3 Skolelærernes oplevelser og erfaringer

I interviewene med skolelærerne, kan der heller ikke her spores systematiske forskelle mellem skoletyper i forhold til den overordnede vurdering af bedømmelsesordningen. Repræsentanter for alle skoletyper udtrykker frustration over, at man ikke får lov at følge eleverne til dørs, og at dialogen med censorerne mangler. Igen handler dialogen både om at sikre rettepålideligheden (ved at få lærernes øjne på bedømmelsen) og om at få den faglige sparring med censor.

Man føler ikke, at man fører dem til døren. Og jeg har oplevet mange fejl i årenes løb, både fra egen og fra censors side. Rettesikkerheden er bedre, når man er to. Dansk er mere kritisk, da der i højere grad ligger en subjektiv vurdering. (Folkeskolelærer, dansk og engelsk)

I matematik er der en ret præcis rettevejledning, som ensretter. Det vigtigste er jo, at de får den rigtige undervisning undervejs, som passer til kravene til faget. Man har brug for at vide, hvad kravene er. (Efterskolelærer, matematik og naturfag)

Jeg savner opfølgning på prøven over for elever, og at man får lov at sikre, at ens elever får den rette karakter. Tilfredsheden med arbejdet er væk, Man får ikke lov at følge dem til dørs. Jeg føler, at det hele kan være lige meget så. (Folkeskolelærer, matematik og naturfag)

Ligesom for censorerne, er der også i lærerinterviewene en tendens til, at det især er i dansk og i sprogfagene, at rettepålideligheden problematiseres. Igen er der dog også variation inden for fagene, hvor der fx både er matematiklærere, der mener, at rettevejledningerne ensretter karaktergivning, mens andre forsat er bekymrede for, om deres elever får den korrekte karakter. På tværs af alle fag er der en overvejende frustration over ikke at være med til at afslutte arbejdet med eleverne, og over ikke at få feedbacken fra censorerne.

3.2 Feedback

En af konsekvenserne af indførelsen af ordningen med én bedømmer er, at kontakten mellem censorer og lærer er fjernet, og dermed også lærernes indbyggede mulighed for at få feedback på deres elevers præstationer. Mens det førhen ikke var pålagt censorer at give egentlig faglig feedback, oplevede flere lærere og censorer, at det indgik som en central del af prøve- og bedømmelsessituationen og som noget, der kunne hjælpe lærerne med at evaluere og forbedre deres undervisning.

I spørgeskemaundersøgelsen spørges censorerne, hvordan de oplever det, at muligheden for at give feedback er forsvundet. Her svarer censorer på tværs af fag nogenlunde ens. Cirka 49 % af danskensorerne, 43 % af matematikcensorerne og 45 % af sprogfagscensorerne oplever det som "problematisk" eller "meget problematisk", at muligheden for feedback er forsvundet. Mellem 17 og 23 % forholder sig neutralt, mens ca. 32 % af både dansk-, matematik- og sprogcensorer oplever det som "uproblematisk" eller "helt uproblematisk", jf. Tabel 3.5.

Tabel 3.5 Censorbesvarelser på spørgsmålet: "Hvordan oplever du det som censor, at muligheden for at give faglig feedback til klassens lærer i forbindelse med bedømmelse af skriftlige opgaver er forsvundet efter indførelsen af ordningen med én bedømmer?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag
Meget problematisk	17,7	15,4	14,0
Problematiske	30,9	27,8	30,7
Hverken problematisk eller uproblematisk	17,4	22,7	19,4
Uproblematisk	21,6	19,4	17,2
Helt uproblematisk	10,6	12,7	15,1
Ved ikke	1,8	2,0	3,8
Antal besvarelser	282	299	186

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,551. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til censorer.

På den ene side opleves det som en fordel af censorer i både matematik, dansk og sprogfag, at man ikke længere skal bruge tid på at forhandle karaktererne på plads. På den anden side udtrykkes der også frustration i fagene over, at man ikke længere kan give feedback til lærerne. Dobbelttheden udtrykkes i følgende citater:

Feedbacken mangler, men rettetarbejdet fungerer okay. Man skal ikke diskutere. Rettevejledningen giver gennemsigthed, og det er en fordel for eleverne. Men det er også rart med dialog. Så er man sikker på at få sagt tingene til lærerne. (Censor i matematik)

Jeg mangler konferencen med faglæreren, hvor jeg kan sige, det her skal du have fokus på i din undervisning for at hæve elevernes generelle karakterer. Omvendt har jeg også oplevet faglærere, der har trukket for højt op taktisk, fordi de ved, jeg giver dumpekarakter. (Censor i dansk)

Det er rart at kunne snakke tingene igennem, men det er også rart med friheden. (Censor i engelsk)

Baseret på de kvalitative indsigter er der altså variation i forhold til, hvor bekymret man er for retteligheder, samt hvordan man vægter friheden til at rette uden indblanding – kontra muligheden for at sparre med lærerne og give dem feedback. Selvom der er censorer, der oplever bedømmelsesarbejdet under den nye ordning som "befriende" og "nemmere", og selvom der er variation i oplevelsen af subjektivitet i rettetarbejdet, er den overordnede holdning i de kvalitative interview dog, at lærerne bør have en form feedback om prøvebesvarelserne. Det giver både mulighed for at opdatere lærerne om den generelle udvikling inden for faget samt mulighed for at give information til den individuelle lærer om, hvor tingene går galt eller lykkes rigtigt godt.

I lighed med censorerne er der også mange skoleledere, der – på tværs af skolekarakteristik – efterspørger dialogen mellem censorerne og skolelærerne. Her er der især fokus på, at lærerne mister slutevalueringen af deres arbejde, at dialogen med censorerne udvikler lærernes faglige kompetencer og indsigt i, hvad deres elever kan og ikke kan, og at lærerne ikke kan give eleverne en tilfredsstillende forklaring på deres karakter. Skolelærerne udtrykker overordnet det samme, og der efterspørges viden om resultatet af egen undervisning.

Jeg er enig i, at det er dejligt ikke at skulle rette, men det overgår ikke frustrationen. Man aner ikke, om ens undervisning har båret frugt. Jeg vidste ikke, at jeg ville savne dialogen, som eksisterede under den gamle ordning, men det gør jeg nu. Der mangler et sikkerhedsnet. Det er respektløst. (Efterskolelærer, dansk og engelsk)

De kvalitative interview viser, at enkelte censorer stadig tager tid til at give en form for uformel, skriftlig feedback til lærerne eller skolerne. Selvom man ikke længere har dialogen med lærerne, er der flere censorer, der sender skriftlig feedback til skolerne sammen med prøvebedømmelserne. Nogle censorer gør det hver gang, hvor de beskriver gode og dårlige elementer på tværs af opgaverne, andre giver feedback i tilfælde, hvor de vurderer, at der er særligt behov for det. Andre censorer antager, at skolen selv tager kontakt, hvis de har spørgsmål til bedømmelserne. Nogle matematikcensorer fortæller endvidere, at de bliver opfordret til at sende diagrammer over den samlede score for klassen for hver opgave i prøvesættet tilbage til skolerne, når de har rettet opgaverne, men at det ikke er et krav. Selvom man sender statistikken og/eller kommentarer til opgaven, opleves det for censorerne som utilfredsstillende, at de ikke ved, om informationen går videre til lærerne. Denne oplevelse af, at man ikke ved, om feedbacken når frem til de respektive lærere, går igen blandt censorer i forskellige fag:

Jeg skriver ned til lederen, hvis der er noget helt galt i opgaverne, men hvis det er småjusteringer, hvor det er kommet bag på læreren, at noget blev ændret i april eller noget, så ville jeg aldrig gå til lederen. (Censor i matematik)

Nogle gange dør den nok også ved lederen – de glemmer det eller fatter det ikke. Man ville være mere ærlig, hvis man kunne gå direkte til læreren. (Censor i dansk)

Hvis censor har haft fokus på noget i sin bedømmelse, som læreren måske ikke har vægtet i sin undervisning, har man [tidligere, red.] snakket sig til rette. Man har endvidere kunnet indgå i en samtale om, hvorvidt der er nogle områder i lærernes undervisning, som er blevet underprioriteret. Det er sværere at tage snakken med skolelederen. (Censor i engelsk)

Overordnet set er der flere censorer på tværs af fag, som ser det som et problem, at man ikke har den mundtlige dialog med lærerne. Der er dog også flere, der ser skriftlig feedback som et alternativ. Som udtrykt i ovenstående citater er udfordringen ved dette fra censorernes perspektiv, at informationen ikke er målrettet lærerne.

Samme tendens gør sig i endnu tydeligere grad gældende blandt skolelederne, hvor mellem knap 38 og 43 % i høj grad vurderer det som problematisk, at lærerne ikke længere får feedback fra censorerne, jf. Tabel 3.6. Dette kan skyldes, at den tidligere ordning (med to bedømmere) til en vis grad sikrede lærerne muligheden for at få feedback – en feedback, som skolelederne nu i højere grad skal sørge for, at lærerne får mulighed for at få fra et andet sted.

Tabel 3.6 Skolelederbesvarelser på spørgsmålet "I hvilken grad vurderer du det som problematisk, at der under den nuværende bedømmelsesordning ikke gives feedback fra censorerne til lærernes undervisning?" Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler
I høj grad	41,5	43,2	37,8
I nogen grad	35,4	30,4	37,8
I mindre grad	17,0	22,4	21,4
Slet ikke	6,1	4,0	3,1
Antal besvarelser	229	125	95

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,624. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til skoleledere.

Skolelederne vil dog også gerne selv have information om generelle tendenser i de enkelte klasser. Nedenstående citat opsummerer skoleledernes overordnede holdning:

Vi har erfaringsudveksling på kommuneniveau. Men det er mere opgavetendenser. Det er jo ikke det samme som en detaljeret feedback. Den eneste form for detaljeret feedback, som vi får, det er, når vores elever klager eller dumper af forskellige årsager. For eksempel på grund af kildehenvisninger. Feedbacken kommer i censorens skriv/begrundelse. Men det ville være fedt med feedback til lærer og skole. Det er ærgerligt, at der er nogle elever, der skal "gokkes i nøden", før vi får feedback og information. (Skoleleder)

3.2.1 Feedbackformer

I spørgeskemaet har skolelederne forholdt sig til, i hvilket omfang de vil bevilge arbejdstid til, at læreren kan arbejde med en række feedbackformer. Tabel 3.7 viser, at skolelederne vil være mest tilbøjelige til at bevilge, at enten hele eller størstedelen af arbejdstiden bliver dækket, hvis lærerne (ved behov) fik mulighed for at tage telefonisk kontakt til censorerne i forhold til feedback. Dernæst vælger flest skoleledere en løsning, der indbefatter et "andet fast format". Der er dog blandt skolelederne overordnet set en stor vilje til at dække enten hele eller det meste af den arbejdstid, lærerne ville skulle bruge på feedback.

Tabel 3.7 Skolelederbesvarelser på spørgsmålet "I hvilken grad vil du bevilge arbejdstid til, at læreren kan arbejde med..." Fordelt på skoletype. Procent.

I hvilken grad vil du bevilge arbejdstid til, at læreren kan arbejde med...	Folkeskoler	Frie grundskoler	Efterskole
<i>... en kort skriftlig redegørelse fra censor i forhold til, hvad censor vurderer bør være fokuspunkter i undervisningen?</i>			
Fuldstændig dækning af arbejdstid	50,7	44,7	49,5
Det meste dækket	20,9	28,5	27,8
En mindre del dækket	15,6	13,0	15,5
Slet ikke	12,9	13,8	7,2
Antal besvarelser	225	123	97
<i>... et afkrydsningsskema fra censor i forhold til, hvad censor vurderer bør være fokuspunkter i undervisningen?</i>			
Fuldstændig dækning af arbejdstid	53,6	46,0	44,3
Det meste dækket	20,5	26,6	28,9
En mindre del dækket	13,0	9,7	17,5
Slet ikke	13,0	17,7	9,3
Antal besvarelser	224	124	97
<i>... mulighed for, at læreren kan tage telefonisk kontakt til censor i de skriftlige prøver i forhold til feedback på undervisningen, hvis der er behov?</i>			
Fuldstændig dækning af arbejdstid	61,1	53,7	53,1
Det meste dækket	19,9	25,2	29,6
En mindre del dækket	11,1	8,9	10,2
Slet ikke	8,0	12,2	7,1
Antal besvarelser	226	123	98
<i>... et andet fast format for tilbagemelding til læreren el. skolen i forhold til fokuspunkter i undervisningen?</i>			
Fuldstændig dækning af arbejdstid	52,7	44,7	41,7
Det meste dækket	21,2	33,3	37,5
En mindre del dækket	17,1	7,3	13,5
Slet ikke	9,0	14,6	7,3
Antal besvarelser	222	123	96

Anm.: Svarfordelingerne er testet med Chi²-test. Kun for spørgsmålet "Et andet fast format for tilbagemelding til læreren el. skolen i forhold til fokuspunkter i undervisningen" er der statistiske signifikante forskelle mellem svarfordelingerne.

Kilde: VIVEs spørgeskema til skoleledere.

Samtidig har vi i interviewene også diskuteret, hvilke andre faste *feedbackformater* der kunne anvendes til skolelederne og lærerne. Det er de færreste censorer, der mener, at en fast skabelon for feedback er en god ide, fordi det bliver for "copy-paste" eller "firkantet". Censorerne har desuden svært ved at forestille sig, hvordan skabelonen skulle udformes, og hvad indholdet skulle være. Der er dog også flere censorer, der har en positiv holdning til et egentlig krav om, at der skal sendes et følgebrev med, når man sender en classes prøveopgaver tilbage.

I Københavns kommune har der tidligere været et forsøg, hvor der blev udarbejdet en rapport på klasseniveau over, hvordan eleverne klarer sig. De skoleledere, der har oplevet forsøget, beskriver det som velfungerende og understreger, hvor vigtigt det er med detaljeret feedback på klasseniveau. Nogle skoleledere peger på, at det var nyttigt med den information, som man tidligere fik fra PEU (prøve, evaluering, undervisning), som opsummerede censorerfaringerne.

Også skolelærerne oplever, at de mangler feedback om elevernes resultater. Lærerne forholder sig kun overfladisk til prøveresultatet, når det foreligger. Langt de fleste lærere fortæller, at de får karakteroversigterne for deres klasse, men ellers ikke sparrer systematisk med nogen om disse. Enkelte af skolelærerne sparrer fortsat med de relevante fagteams på skolen, mens andre fortæller, at muligheden for sparring med kollegaer netop har ændret sig:

Hvis man havde en opgave, hvor man fx lå to karakterer fra hinanden, ville man bede en af de andre om at læse førhen. Det gør man jo ikke mere. (Lærer i dansk og matematik)

Der er også forskel på, om lærerne læser elevernes opgaver igennem. Nogle gør det, hvis de oplever, at der er noget, der "stikker ud". Andre har en lidt anden tilgang:

Jeg kigger på resultaterne og siger "nå, det var ikke så godt". Men jeg kan jo ikke vide noget eller forklare eleverne noget, for jeg har jo ikke rettet. (Lærer i dansk og tysk)

Blandt skolelærerne er der også forskellige holdninger til, hvorvidt skriftlig feedback er gavnligt, samt til hvilken form feedbacken skal have. Mange af lærerne – især i dansk og i sprogfagene – ser den mundtlige dialog som det optimale, men de understreger også, at skriftlig feedback er at foretrække frem for ingenting:

Det er en anden snak, man får mundtlig frem for skriftlig. Det virker mere positivt, og tonen er en anden. Men hellere skriftlig end ingen feedback. (Lærer i dansk)

På tværs af fagene er der flere lærere, der gerne vil have generel information om eleverne i deres klasse, men der er uenighed om, hvorvidt feedbacken bør gives i en fast skabelon. Dette gælder også inden for fagene, hvor der fx er lærere i både dansk og engelsk, som er mere positive over for afkrydsningsskemaer, mens andre mener, at feedbacken i disse fag bør være mere åben:

Feedbacken må ikke være et afkrydsningsskema, for det kan tolkes på mange forskellige måder. Det kan være meget forskelligt, hvad der vægtes. (Lærer i engelsk)

3.3 Delkonklusion

Den overordnede oplevelse af bedømmelsesordningen varierer både mellem og inden for grupperne af censorer, skoleledere og skolelærere. Mens der er mellem 50 og 60 % af censorerne, der oplever, at ordningen med én bedømmer fungerer godt eller særdeles godt, er dette kun tilfældet for mellem 22 og 33 % af skolelederne. Inden for gruppen af censorer er matematikcensorerne generelt mere positive over for bedømmelsesordningen end dansk- og sprogfagscensorerne. Indsigterne fra de kvalitative interview peger på, at der er en større oplevet subjektivitet i karaktergivningen i dansk og sprog end i matematik, og at denne oplevede subjektivitet kan give anledning til bekymring for rettespørgsmålet. Det gælder også hos skolelærerne, hvor der også udtrykkes størst bekymring for rettespørgsmålet blandt dansklærerne og sprogfagslærerne. Der er dog også både matematikcensorer og matematiklærere, der udtrykker bekymring for rettespørgsmålet. For de skoleledere, der har været ledere både før og efter indførelsen af ordningen, er langt de fleste ledere blevet mere usikre på bedømmelserne. For mange af skolelederne forstærkes problematikken af, at karaktererne i dag betyder mere for elevernes adgang til ungdomsuddannelserne. Især skoleledere fra skoler med lave eller middel karakterniveauer fremhæver desuden, at det er et problem, at det i højere grad udelukkende er elevernes og forældrenes opgave at klage over karaktererne, da lærerne ikke længere læser opgaverne, og derfor er det sværere for dem at rådgive om

klager. Oplevelsen er nemlig, at det især er elever i den lave ende, som undlader at klage – selvom det specielt er dem, der kan få problemer med at komme videre i uddannelsessystemet.

Både censorerne, skolelederne og skolelærerne udtrykker frustration over, at skolelærerne ikke længere får feedback fra censorerne. Selvom der på censorsiden også udtrykkes fordele ved ikke at skulle forhandle med lærerne, og selvom skolelederne også kan se fordelene ved, at lærerne ikke skal have tid til at rette, er den overordnede oplevelse, at feedbacken har stor værdi. Flere censorer giver i dag alligevel en eller anden form for skriftlig feedback på opgaverne – det er dog uklart, om denne feedback havner hos skolens ledelse eller når ud til lærerne. Langt de fleste lærere fortæller, at de ikke systematisk forholder sig til prøveresultatet, ikke får systematisk feedback eller sparring på elevernes resultater, men at de blot ser en oversigt over karaktererne. Nogle lærere læser de opgaver, som ”stikker ud”.

Der er forskellige holdninger til skriftlig feedback samt forskellige forslag til feedbackformat. Dog foretrækker lærerne imidlertid også den mundtlige, faglige feedback. Skolelederne udviser i et vist omfang villighed til at dække enten hele eller det meste af den arbejdstid, lærerne ville skulle bruge i forbindelse med at få feedback, Lederne er da også mest tilbøjelige til at ville dække arbejdstid, hvis feedbacken foregår telefonisk og efter behov. Der er også en positiv indstilling til andre faste feedbackformater.

4 Censorernes arbejdsvilkår og kvalificering af deres arbejde

I dette kapitel ser vi nærmere på censorernes arbejdsvilkår samt på kvalificering af deres arbejde. Mere specifikt analyseres censorernes og skoleledernes vurdering af censorrollen gennem belysning af årsagerne til at (fra)vælge hvervet som censorer, samt på skoleledernes oplevelse af at have beskikkede censorer på skolen. Da Undervisningsministeriet beskikker censorer efter indstilling fra skolernes ledere, afdækkes, hvad skolelederne gør for at finde og kvalificere lærere til hvervet. Endelig undersøges, hvordan censorerne oplever og benytter sig af de eksisterende bedømmelsesværktøjer, censorkurser og rettegrupper.

Konkret undersøges spørgsmålene:

- Hvordan vurderer censorerne og skolelederne og lærerne censorrollen?
- Hvad gør skolelederne for at finde og kvalificere censorerne til censorhvervet?
- Hvordan samarbejder censorer og skoler om bedømmelserne, herunder også i forhold til klager?
- Hvordan oplever/benytter censorerne de eksisterende bedømmelsesværktøjer, censorkurser og rettegrupper?

4.1 Vurderingen af censorrollen

På tværs af alle fag fortæller censorerne i interviewene, at de har valgt censorhvervet, fordi det giver en faglig opdatering og kontakt til de faglige miljøer, hvor man blandt andet kan sparre med andre censorkolleger. Mere specifikt giver censorhvervet mulighed for at følge med i, hvordan faget udvikler sig, og der er en oplevelse af, at man kommer ind i faget på en anden måde, som kommer både egne elever og skolen til gavn. Flere censorer fortæller endvidere, at de agerer som sparringspartner for faglærerne på skolen, og at de bliver mere målrettet i forhold til egen undervisning.

Skolelederne er overvejende positive over at have censorer på skolen. Tabel 4.1 illustrerer en umiddelbar tendens til, at skoleledere på folkeskoler er mere positivt stemte end skoleledere på henholdsvis frie og private grundskoler og efterskoler. Efterskolelederne oplever sammenlignet med de andre skoletyper, at beskikkede censorer i lærerstaben giver mindst værdi. Ligesom for den overordnede oplevelse af bedømmelsesordningen kan forskellen mellem forskellige skoletyper ikke spores i interviewene med skolelederne.

Tabel 4.1 Skolelederbesvarelser på spørgsmålet "I hvilken grad vurderer du, at det giver skolen værdi, at en eller flere lærere er beskikkede som censorer i de skriftlige prøver?" Fordelt på skoletype. Procent.

	Folkeskole	Frie grundskoler	Efterskoler
I høj grad	50,2	38,9	37,8
I nogen grad	41,5	44,4	37,8
I mindre grad	8,3	14,3	18,4
Slet ikke	0,0	2,4	6,1
Antal skoler	229	126	98

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,001. Dermed er der statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til skoleledere.

Kun et fåtal af skolelederne udtrykker overvejende ulemper ved at have censorer på skolen, og her er både en repræsentant fra en efterskole, en folkeskole og en friskole. Disse ledere fokuserer på, at censorerne koster ressourcer, og at der ikke er sikkerhed for, at censorarbejdet ikke går ud over lærernes almindelige arbejde. For de mere positive skoleledere ses censorhvervet både som en kompetenceudvikling af de enkelte lærere, og som en ressource for skolen som helhed. I lighed med censorerne selv ser skolelederne nemlig censorerne som nogle, man kan trække på i forhold til at indhente viden om retningslinjerne inden for faget, ligesom censorerne også kan bruges som sparringspartnere for faggruppen.

Selvom skolelederne generelt oplever, at det kan give skolen værdi at have beskikkede censorer i lærerstaben, så oplever skolelederne dog også (mellem 48 og 60 %), at lærerne kun i mindre grad selv er interesserede i at besidde hvervet. Dette ses i Tabel 4.2.

Tabel 4.2 Skolelederbesvarelser på spørgsmålet: "I hvilken grad vurderer du som skoleleder, at lærerne på din skole er interesserede i at blive beskikkede som censorer i de skriftlige prøver i 9. eller 10. klasse?" Fordelt på skoletype. Procent.

	Folkeskole	Frie grundskoler	Efterskoler
I høj grad	3,5	3,2	4,1
I nogen grad	25,7	25,6	19,4
I mindre grad	59,6	48,0	54,1
Slet ikke	11,3	23,2	22,4
Antal skoler	230	125	98

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,056. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til skoleledere.

I spørgeskemaundersøgelsen er skolelederne spurgt om deres vurdering af, hvorfor lærerne ikke er interesserede i censorhvervet (Bilagstabel 2.6). Godt halvdelen af de skoleledere, der oplever, at lærere i staben kun i mindre grad eller slet ikke er interesserede i at blive censorer, vurderer, at det i særlig høj grad skyldes, at det er for tidskrævende at være censor. Lidt under halvdelen vurderer, at det skyldes, at det er besværligt at være censor, og at betalingen ikke er tilstrækkelig. Årsager til fravalg af censorhvervet diskuteres også i interviewene. På tværs af både censorer, skoleledere og skolelærere er opfattelsen, at mange lærer vælger censorhvervet fra, fordi arbejdet ligger på et tidspunkt af året, hvor man i forvejen har travlt. Det koster på den måde også noget i forhold til familien og muligheden for at nyde foråret. Desuden er der mange, der ikke ser det som en spændende opgave at rette opgaver. Pengene ses af flertallet af censorerne som en god compensation, om end det kun er et fåtal af censorerne, der nævner pengene som den vigtigste årsag til hvervet. For mange

af lærerne opvejer pengene heller ikke de nævnte ulemper. Der er desuden flere af lærerne, som også har en opfattelse af, at det er svært at nå at rette mængden af opgaver på den normerede tid. Mere tid per opgave eller færre opgaver ville gøre hvervet mere attraktivt. Muligheden for at varetage en halv censorstilling, hvor man retter mindre, nævnes også eksplicit af én af censorerne som årsagen til, at jobbet kan varetages ved siden af den almindelige undervisning.

I flere af fokusgruppeinterviewene diskuteres det, hvorvidt det ville være en fordel, at arbejdet blev del af årsnormen i det almindelige lærerjob, så man blev fritaget fra andre opgaver, mens man rettede opgaver. For nogle lærere ville det gøre hvervet mere interessant, mens det for andre lærere ikke ville give mening, fordi de hvert år har ansvaret for at føre elever op til prøve i netop den periode, hvor rettearbejdet skal varetages – eller fordi opgaven netop ikke er en opgave, de er interesseret i. Endelig er der også en tendens til, at lederens aktive arbejde for at finde beskikkede censorer gør en forskel. For eksempel fortæller en lærer på en skole, hvor der ikke er beskikkede censorer, at der egentlig er flere lærere, der godt kunne tænke sig hvervet, men at hvervet ikke bliver set som en ressource af lederen. Ligeledes fortæller en anden lærer, at der tidligere har været beskikkede censorer på lærerens skole, men at man har fået ny ledelse, som ikke er så aktiv, så derfor er der ikke længere nogen censorer. Flere af de beskikkede censorer, som har påtaget sig hvervet, har gjort dette, fordi de blev opfordret af deres leder, som netop "talte hvervet op". I det efterfølgende afsnit afdækkes skoleledernes arbejde med at finde og kvalificere lærere til at blive beskikkede censorer.

4.2 Skoleledernes arbejde for at finde nye censorer og kvalificere deres arbejde

Som det ses i Tabel 4.3 nedenfor, er der mellem 40 og 51 % af skolelederne, som siden maj 2016 har opfordret en eller flere af lærerne på deres skole til at blive beskikket censor. I de kvalitative interview er det især ledere, som ser censorhvervet som en fordel for lærerne og skolen, som aktivt opfordrer deres lærere til at påtage sig hvervet. På nogle skoler er det dog ikke nødvendigt med en opfordring, fordi lærerne påtager sig hvervet på eget initiativ. De ledere, som ikke ser fordele ved at have beskikkede censorer på skolen, opfordrer ikke aktivt til hvervet, ligesom der også er ledere, der ikke har overvejet, om det faktisk er deres opgave at motivere lærerne til jobbet. For at øge interessen for censorhvervet er et forslag fra en skoleleder at afholde "før-censorkurser", hvor lærerne kunne få et større indblik i, hvad rollen som beskikket censor egentlig indebærer.

Tabel 4.3 Skolelederbesvarelser på spørgsmålet "Har du som skoleleder i perioden siden maj 2016 opfordret en eller flere lærere på din skole til at blive indstillede til beskikkelse som censorer i skriftlige prøver i 9. eller 10. klasse?" Fordelt på skoletype. Procent.

	Folkeskole	Frie grundskoler	Efterskoler
Ja	10,4	13,5	25,5
Ja, nogle få	35,1	27,0	25,5
Nej, ingen	54,6	59,5	49,0
Antal skoler	132	126	98

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,006. Dermed er der statistisk signifikante forskelle på svarfordelingerne. Svarkategorierne "Ja, mange" og "Ja, nogle" er slået sammen til kategorien "Ja".

Kilde: VIVEs spørgeskema til skoleledere.

I interviewene med skolelederne er det også forskelligt, hvorvidt lederne gør sig overvejelser om, *hvem* de opfordrer til at være beskikket censor. Én skoleleder fortæller fx, at det skal være de lærere, der er fagligt og personligt kompetente til hvervet, og som kan håndtere klagesager. For mange af

lederne er holdningen dog, at hvis lærerne opfylder de formelle krav, så er de kvalificerede – for man har jo kun kvalificerede lærere ansat på skolen. Hvis disse ledere opfordrer til censorhvervet, er opfordringen derfor til alle lærere – fx på et lærermøde. I forhold til den løbende kvalificering af arbejdet, er det også forskelligt, hvorvidt skolelederne følger med i, om lærerne tager på censorkurser, og hvorvidt man som leder egentlig er klar over, om kurserne er obligatoriske for censorerne eller ej. En skoleleder fortæller:

På den ene side går ministeriet igennem os for at få lærerne, men derefter kobles vi helt af – så kan det jo ikke være mit ansvar, at de bliver godt kvalificeret. Jeg bliver ikke orienteret af ministeriet om kurser og så videre for censorerne – det er dem selv, der orienterer mig, når de vil afsted. (Skoleleder)

Selvom flere ledere godt kan se vigtigheden af at have et kvalificeret censorkorps, pointerer langt de fleste også, at de beskikkede censorer jo er ansat af ministeriet, og derfor må det være ministeriets ansvar at sikre kvaliteten. Hvis opgaven skal ligge på skolen, er den overordnede holdning, at opgaven skal understøttes af ressourcer.

4.3 Samarbejdet mellem censorer og skolerne

Dette afsnit beskriver censorer og skolelederes perspektiver på samarbejdet om bedømmelsesprocessen, herunder også samarbejdet i forbindelse med klager. Samarbejdet mellem censorerne og skolerne fungerer grundlæggende i to separate spor. Det ene spor er af praktisk og logistisk karakter, og omhandler eksempelvis forsendelse eller scan af opgaver, print og kopier af opgaver og håndtering af kontaktoplysninger og anonymisering. Det andet spor omhandler samarbejdet og procedurerne i forbindelse med klager. Klageproceduren er ikke formelt ændret med den nye ordning, men elevens lærer har med den nye ordning ikke indsigt i elevens besvarelse af den skriftlige prøve. Læreren har derfor sværere ved at vejlede eleverne i forbindelse med overvejelser om klager.

Den overordnede konklusion fra de kvalitative interview er, at der generelt ikke er tale om et omfangsrigt samarbejde, og at det fungerer godt med minimal kontakt. Der er således ikke et ønske om, at der skal være yderligere kontakt mellem censorer og skoler. Der er ikke identificeret forskelle på tværs af skoletyper eller fag.

Når det kører, er der ikke noget samarbejde mellem skolerne og censoren. (Censor i tysk)

4.3.1 Praktisk samarbejde

I forhold til det første spor, det praktiske samarbejde, vægter censorerne samarbejdet med skolesekretærene højt. Skolesekretærene beskrives som kompetente og samarbejdsvillige i forbindelse med håndtering af forsendelse af opgaver mv. En del censorer har dog oplevet eksempler på, at der sker fejl. Dette lader dog ikke til at være typisk for samarbejdet, men at det dog er frustrerende, når det sker. Det kunne eksempelvis være, at opgaverne ikke er blevet anonymiseret i tilstrækkelig grad, eller at der decideret mangler sider fra en eller flere opgaver. For at komme disse udfordringer i forkøbet benytter flere censorer sig af muligheden for at sende et skriv om formalia, proces og ønsker til skolerne forud for bedømmelsesperioden. Dette opleves som en hjælp for både censorer og skolerne.

Censorerne er blevet gode til at sende et indledende skriv, før bedømmelserne går i gang. Men det kunne være nemmere med et standardiseret skriv. Det er godt lige at få reminder, om hvad der skal være styr på. (Skoleleder)

Størstedelen af censorerne ønsker desuden, at de modtager en kopi af originalopgaven, så det er nemmere at skrive kommentarer og bruge 'den røde pen' i forbindelse med bedømmelsen. Her er censorernes erfaringer meget forskellige fra skole til skole. Nogle skoler er meget behjælpelige med kopiering eller scanning af opgaver, mens andre ikke er. Flere censorer oplever også, at de bruger for lang tid på at finde kontaktoplysninger på de relevante medarbejdere på de skoler, som de er censor for. De har som oftest et telefonnummer, men har i højere grad behov for en mailadresse, der gør det muligt at tage kontakt til skolerne uden for almindelig arbejdstid. På nuværende tidspunkt bruger flere censorer lang tid på at søge efter kontaktoplysninger til de enkelte skoler.

4.3.2 Klager

I forhold til det andet spor, klagesager, er oplevelsen, at der generelt er et gnidningsfrit samarbejde mellem censorer og skoler. Klageproceduren er som nævnt ikke ændret formelt, da det stadig er forældrenes og elevens ansvar at klage over en bedømmelse. Forskellen ligger i, at lærerne tidligere kunne opfordre eleven til at klage på baggrund af den samtale, læreren har haft om opgaven med censor. I dag har læreren ikke indsigt i elevens besvarelse af prøven. Klagen tages op med skolelederen, som vælger, hvorvidt klagen skal forfølges. Her har skolelederen i første omgang mulighed for at kontakte censor med henblik på at indhente en udtalelse omkring bedømmelsen. Hvis der fortsat er uenighed omkring bedømmelsen, kan skoleledelsen indstille til ombødelse gennem STUK.

Det gnidningsfrie samarbejde skyldes, dels at der er relativt få klagesager, jf. Tabel 4.4, der viser, hvor ofte censorerne oplever, at de bliver kontaktet af skolerne med henblik på at få en udtalelse i forbindelse med en klagesag. For matematik og sprogfag har hhv. 11 og 16 % oplevet at blive kontaktet, mens de resterende ikke er blevet kontaktet med henblik på at få en udtalelse. Billedet er lidt anderledes i dansk, hvor 43 % af censorerne har oplevet at blive kontaktet på mindst én af de skoler, hvor de har ageret censor. Der tegner sig således et billede af, at der generelt er få klagesager, og at når der er klager, så er det primært i forbindelse med bedømmelse af danskopgaver.

Tabel 4.4 Censorbesvarelser på spørgsmålet "Hvor ofte oplever du, at en skole har kontaktet dig med henblik på at indhente en udtalelse i forbindelse med en klagesag?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag
På ingen af de skoler, hvor jeg har været censor	57,5	88,9	84,0
På 1-25 % af de skoler, hvor jeg har været censor	40,4	11,1	16,0
Flere end 25 % af skolerne	2,1	0,0	0,0
Antal besvarelser	282	298	187

Anm.: Svarfordelingerne er testet med Chi2: p=0,000. Dermed er der statistisk signifikante forskelle på svarfordelingerne
Kilde: VIVEs spørgeskema til censorer.

Det gnidningsfrie samarbejde afspejler desuden også at, at når der er klagesager, så følges de faste procedurer og formalia, hvilket opleves uproblematisk.

Jeg bruger klagevejledningen. Jeg gemmer mine kommentarer og tjekker opgaven igennem igen. Så svarer jeg, hvad der ligger til grund for den karakter, jeg har givet. Kommunikationen [med skolen, red.] har været fin og gik gennem skolelederen. [Problemet, red.]

var noget med kilder, der ikke var til diskussion. Så sagde jeg til lederen, at han skulle opdatere sine lærere på kravene, hvis det var en generel ting på skolen. (Censor i dansk)

De skoleledere, der har erfaring med klagesager, giver i de kvalitative interview udtryk for, at de i højere grad end tidligere har kontakt med de elever og forældre, som ønsker at klage. Skolelederne giver udtryk for, at de fleste forældre godt kan udarbejde klagen, men at nogle har sværere ved det end andre. Skolelederne oplever selv, at de er klædt på til at håndtere denne kommunikation. Bilagstabel 2.14 og Bilagstabel 2.15 viser, at 89 % af censorerne ved, hvordan de skal agere, når en skole klager over en bedømmelse, og at tilsvarende 89 % af skolelederne oplever, at dialogen med censor i forbindelse med klager i nogen eller høj grad fungerer tilfredsstillende. De kvalitative interview viser, at censorerne i forbindelse med klagehenvendelser typisk oplever, at deres udtalelse til skolen accepteres, og at den oprindelige karakter sjældent ændres, hvis opgaven bliver ombedømt. Skoleledernes besvarelser på spørgsmål om udtalelser i spørgeskemaundersøgelsen viser ligeledes, at 69 % i høj eller nogen grad oplever, at censorens udtalelser er en tilfredsstillende begrundelse (jf. Bilagstabel 2.16). Der er dog også eksempler på, at censoren enten har fejlbedømt en opgave eller lavet en fejl i forbindelse med indtastningen af karakterer.

Selvom samarbejdet overordnet fungerer godt, er der dog flere, både censorer og skoleledere, der har haft uhensigtsmæssige oplevelser vedrørende samarbejdet i forbindelse med klagesager. Problemerne knytter sig i de tilfælde oftest til et dårligt samarbejde mellem enkeltpersoner, og ikke som generelle problemer med bedømmerordningen eller censorkorpset generelt. Enkelte censorer beskriver, at deres dialog med konkrete skoleledere er problematisk, og at de føler at skolelederen enten 'gemmer sig' i drøftelsen, som reelt i stedet så foregår med en lærer eller forælder, eller at skolelederen har sin egen agenda og derfor ikke er lydhør over for censorens argumenter.

Når det er skolerne, der har en dårlig oplevelse i forbindelse med en klagesag, retter problematikken sig ligeledes mod en bestemt censor og ikke imod selve ordningen. Her kan dialogen med censor og dennes begrundelser ligeledes fremstå mangelfuld, og den videre proces, hvor STUK skal inddrages, kan også opleves som besværlig og langsommelig. Det skyldes blandt andet, at der er et stærkt ønske om at få elevernes endelige karakterer klar, så de kan få deres prøvebeviser. Flere skoleledere beskriver, at de i høj grad er glade for at have beskikkede censorer tilknyttet deres skole i forbindelse med klagesager, da de så kan bruge deres viden og erfaring i forhold til eventuelt selv at lave en ny vurdering af en opgave.

Blandt censorerne er der dog en klar oplevelse af, at de forsøger at hjælpe eleverne og ikke være unødigt strenge, da de godt ved, at karaktererne har stor betydning for elevens videre uddannelse.

[Det er, red.] også fordi, at det er blevet så vigtigt i forhold til deres videre færd i uddannelsessystemet. Det handler for nogle om, om de kommer på gymnasiet eller ej. Så er det os, som forpurrer deres drøm. Men nu, hvor det er blevet så vigtigt, så er det ikke lige dér, at man skal lære dem at tage sig sammen. (Censor i engelsk)

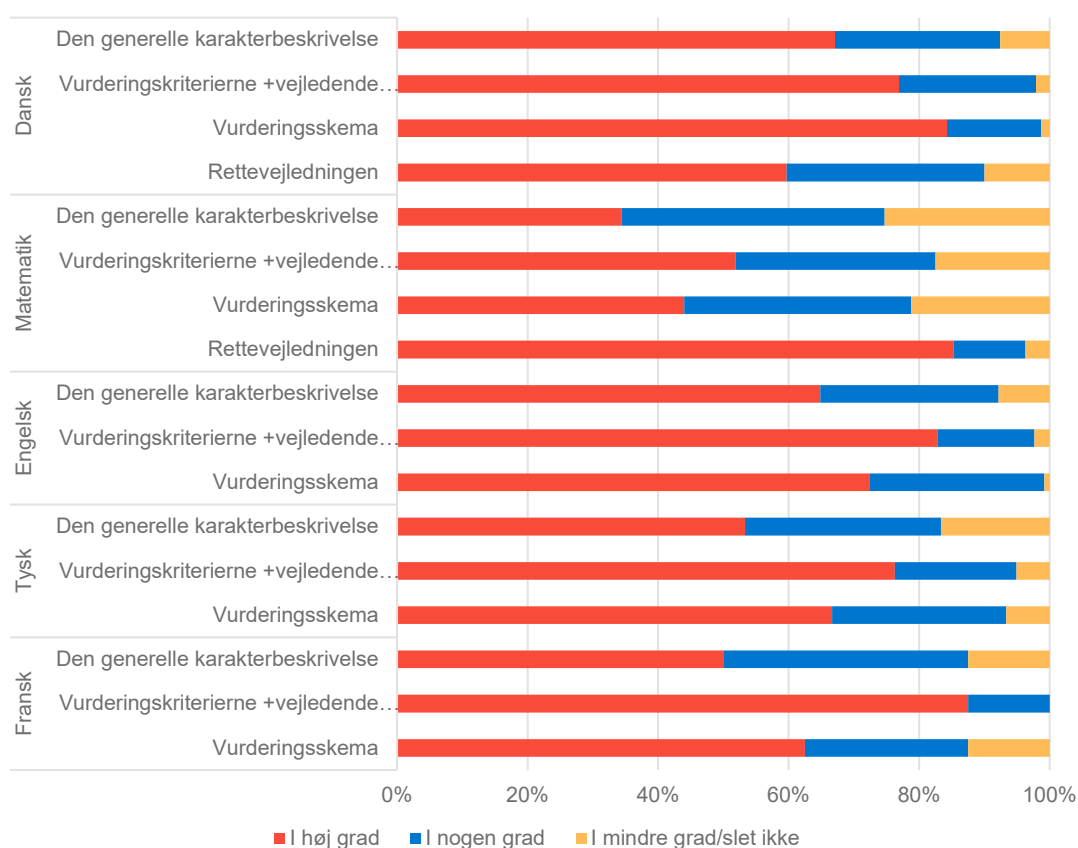
4.4 Bedømmelsesværktøjer

Et centralt element i evalueringen af ordningen med én bedømmer er, om censorerne oplever at kunne gennemføre kvalificerede bedømmelser. Det handler dels om, hvorvidt censorerne oplever at have de fornødne bedømmelsesværktøjer til at foretage en pålidelig og sikker vurdering, dels om censorerne er forberedte på, *hvordan* de skal bedømme.

Censorerens anvendelsesgrad af en række af de mest udbredte bedømmelsesværktøjer (Danmarks Evalueringsinstitut, 2018) fremgår af nedenstående Figur 4.1. Figuren viser, i hvilket omfang censorerne i de fem forskellige fag anvender følgende redskaber:

- Den generelle karakterbeskrivelse for 7-trinsskalaen
- Vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen
- Vurderingsskema, der tager udgangspunkt i den vejledende karakterbeskrivelse⁴
- Rettevejledningen for fagene.⁵

Figur 4.1 Censorbesvarelser på spørgsmålet: "I hvilken grad gør du brug af følgende bedømmelsesværktøjer, når du afgiver karakter i den skriftlige prøve i faget?" Fordelt på fag. Procent.



Kilde: VIVEs spørgeskema til censorer.

Figur 4.1 viser, at vurderingsskemaet, der tager udgangspunkt i den vejledende karakterbeskrivelse er det mest anvendte redskab blandt censorerne i dansk. For matematikcensorer er det klart den (udvidede) rettevejledning, der anvendes. Når det kommer til bedømmelsesværktøjerne blandt censorerne i sprogfagene anvender engelskcensorerne, tyskcensorerne og franskcensorerne i høj grad vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen.

⁴ I matematik er der ikke et vurderingsskema men til gengæld et retteark.

⁵ I sprogfagene er der ikke en rettevejledning

Vi har endvidere bedt censorerne vurdere de fire redskaber i forhold til, om de oplever, at redskabet:

- giver et solidt udgangspunkt for bedømmelsen
- med fordel kunne uddybes
- med fordel kunne forenkles
- er med til at sikre en ensartet bedømmelsespraksis
- sikrer sammenhæng mellem fælles mål/faglige mål/målbeskrivelser og bedømmelsen.

Bilagstabel 2.2-Bilagstabel 2.5 viser de fulde svarfordelinger på disse spørgsmål fordelt på fagtyper. I det følgende giver vi et overblik over censorernes oplevelser med de mest udbredte redskaber.

Hvad angår danskcensorerne, fremhæver de, at vurderingsskemaet giver et solidt udgangspunkt for bedømmelsen. Der er blandende holdninger til, hvorvidt vurderingsskemaet bør uddybes, idet 35 % er enige i, at det bør, mens 42 % er hverken enige eller uenige og 23 % er uenige. Dog er de fleste danskensorer enige i, at vurderingsskemaet er med til at sikre en ensartet bedømmelsespraksis, og at det sikrer sammenhæng mellem mål og bedømmelsen.

Danskensorerne fortæller i de kvalitative interview, at de redskaber, der konkret knytter sig til den enkelte skriftlige prøve, dvs. vurderingsskemaet og rettevejledningen, er gode og uundværlige. De er med til at sikre, at der blandt censorerne lægges vægt på de rigtige ting i den enkelte elevs opgave. Og eksempelvis rettevejledningen og vurderingsskema gør det muligt for censoren at vende tilbage til en opgave, genvurdere eller argumentere for en given bedømmelse. Disse redskaber bidrager således til en systematik i bedømmelserne, som censorerne sætter pris på. De to redskaber står dog sjældent alene, og der er generelt enighed blandt censorerne om, at man også bruger andre redskaber så som karakterbeskrivelserne samt censorens egen fornemmelse og erfaring i forbindelse med vurderingen.

Rettevejledningen sørger for, at subjektiviteten gøres konkret i forhold til, hvad det er, man skal ind og måle på. Rettevejledningen er god til eksempelvis at afklare opgavens disponering i forhold til opgaven eller genrekrav eller brug af kilder. (Censor i dansk)

Jeg bruger vejledningen for prøverne i dansk. Den bruger jeg rigtig meget. Og så bruger jeg ordlyden på karakterbeskrivelserne. Den er super god. Hvis jeg er i tvivl, så læser jeg ordene, og det hjælper. [...] Så kommer man ud over synsningerne. (Censor i dansk)

Censorerne vurderer ikke, at den øgede systematik går ud over den mere holistiske vurdering af opgaven eller prøven.

Blandt matematikcensorerne er det mest udbredte redskab rettevejledningen for faget. På baggrund af tidlige bedømmelser af prøvematerialet udvikler Styrelsen for Undervisning og Kvalitet (STUK) en rettevejledning til matematikcensorer, som giver et indblik i, hvordan forskellige typer af besvarelser bør vægtes. Det udvidede rettevejledningsformat opleves derfor som særdeles godt i forhold til at give et godt udgangspunkt for bedømmelserne (98 % af matematiklærerne er enige i dette). Tilsvarende er 94 % af matematikcensorerne enige i, at rettevejledningen er med til at sikre en ensartet bedømmelsespraksis.

De kvalitative interview understøtter resultaterne, der også ses i tabellerne, som viser, at matematikcensorerne i meget høj grad anvender rettevejledningen, og at den vurderes til at være god og anvendelige. Det samme er også meget tydeligt i de kvalitative interview, hvor der er nærmest enstemmighed om, at rettevejledningen er uundværlig i forhold til at sikre ens og korrekt bedømmelse

på tværs af elevernes opgaver. Det er med andre ord et stort hjælperedskab for matematikcensorerne. Der udtrykkes ligeledes stor tilfredshed med excelrettearket, som både hjælper med en systematisk gennemgang samt karakterforslag og påpeger grænsetilfælde ved karaktergivningen, så man kan undersøge opgaven nærmere.

Rettevejledningen er fantastisk. Man kan godt lave en opgave på flere måder, og rettevejledningen sørger for, at man får det hele med. Jeg laver altid opgaven selv og laver flere løsningsmuligheder. Jeg starter med selv at regne – og gør mig mine egne overvejelser. Så tjekker jeg det af med rettevejledningen. Jeg bruger SkoleKom for at tjekke, om andre har fået andre ting frem. Det er nok med rettevejledningen [som retteværktøj, red.]. (Censor i matematik)

Fælles for de tre sprogfag er, at de alle tre anvender vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen som det primære bedømmelsesværktøj. Generelt er der blandt censorer i alle tre fag en enighed om, at vurderingskriterierne giver et solidt udgangspunkt for bedømmelsen (94 % af engelskcensorerne er enige, mens det samme gælder for alle censorer i fransk og 82 % af censorerne i tysk). Også på tværs af fagene synes der at være enighed om, at kriterierne ikke nødvendigvis bør forenkles, mens der er blandede holdninger til, hvorvidt de med fordel kunne uddybes. Her fordeler besvarelsene sig jævnt henover svarkategorierne (besvarelsene fra fransk-censorer er fordelt ligeligt mellem enige og uenige). Der er på tværs af sprogfagene en oplevelse af, at kriterierne i høj grad er med til at sikre en ensartet bedømmelse, og at de sikrer sammenhæng mellem mål og bedømmelser.

Censorerne i sprogfagene er i de kvalitative interview også overvejende positive i forhold til de uddybede karakterbeskrivelser, om end de er mindre positive end censorerne i dansk og matematik. De vurderer, at redskaberne er blevet bedre og nemmere at fortolke, end de var tidligere, og at de bruges som en art tjekliste for censoren selv for at sikre, at de kommer hele vejen rundt.

Nu kan man bare ikke blive lige så uenig som tidligere. Før havde man nemlig ikke et ark, man kunne støtte sig op ad eller henvise til. Der havde man kun sin egen erfaring. Før kunne man diskutere. Nu er det penslet ud, hvad det kræver at få et 7-tal. (Censor i fransk)

Sprog-censorerne vurderer også, at værktøjerne er med til at give en større grad af synlig objektivitet og gennemsigtighed, hvilket igen er med til at give både eleverne og censorerne højere grad af sikkerhed, eksempelvis i forbindelse med klagesager.

4.5 Rettegrupper og sparring

I forbindelse med behandlingen af lovforslaget om at ændre bedømmelsesordningen til kun at indebære én bedømmer blev der udtrykt bekymring for, om én censor sikrede en tilstrækkelig sikker bedømmelse sammenlignet med den tidligere ordning. Derfor blev der også indført en ordning med rettegrupper, hvor op til fem censorer blev sat sammen, og hvor hver censor har mulighed for faglig sparring med mindst én anden censor om de konkrete opgaver. Den anden censor har mulighed for at læse besvarelsen, når det er nødvendigt, for at sikre en kvalificeret dialog. Censorerne har derudover i gruppen løbende mulighed for at diskutere de faglige spørgsmål, der måtte dukke op i forbindelse med censorarbejdet.

Følgegruppens rapport konkluderede, at kun 81 ud af 361 censorer, svarende til 22 %, havde kontaktet andre censorer for at sparre. Blandt de censorer, der har anvendt rettegrupperne, var holdningen, at rettegrupperne var en god idé. Nærværende evaluering undersøger både, om censorerne har haft kontakt til andre i rettegruppen med henblik på at sparre, samt hvilke barrierer der er for at bruge grupperne, og hvilke alternative sparringsmuligheder de anvender på tværs af fag. Nedenstående Tabel 4.5 bekræfter følgegruppens konklusion, hvor ca. 76-86 % af alle censorer ikke har haft kontakt med rettegruppen. Der er en tendens til, at censorer i dansk i lidt højere grad har gjort brug af rettegrupperne end censorer i sprogfagene, som igen havde lidt mere kontakt end matematikcensorerne.

Tabel 4.5 Censorbesvarelser på spørgsmålet ”Du har haft mulighed for at kontakte en eller flere censorer fra rettegruppen for at diskutere din bedømmelse af en eller flere besvarelser - Kontaktede du nogen fra rettegruppen for at sparre?” Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
Ja, jeg havde kontakt til én anden fra rettegruppen	22,1	12,0	16,7	16,8
Ja, jeg havde kontakt med flere fra rettegruppen	2,1	2,3	5,4	3,0
Nej, jeg havde ikke kontakt med nogen fra rettegruppen	75,8	85,6	78,0	80,2
Antal besvarelser	281	299	186	766

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,004. Dermed er der statistisk signifikante forskelle på svarfordelingerne.
Kilde: VIVEs spørgeskema til censorer.

Af censorerne havde kun ca. 15 % kontakt med én eller flere fra rettegruppen *forud* for bedømmelserne (jf. Bilagstabel 2.6). Dette vidner om, at de fleste først benytter sig af tilbuddet om rettegrupper i tvivlstilfælde. Dette underbygges blandt andet af censorernes besvarelser på, hvorfor de ikke benyttede sig af rettegrupperne. Bilagstabel 2.7 viser, at størstedelen af besvarelserne fordelte sig på svarkategorien ”Der var ikke behov for sparring” (42 % dansk, 44 % matematik, 57 % sprogfag). Imidlertid viser besvarelserne på dette spørgsmål også, at censorerne i høj grad anvender andre sparringsmuligheder end rettegrupperne (41 % dansk, 53 % matematik, 37 % sprogfag).

De kvalitative interview viser, i tråd med ovenstående resultater, at relativt få af censorerne har brugt rettegrupperne. Samtidig viser interviewene, at der ikke efterspørges andre sparringsmuligheder end dem, som censorerne allerede benytter. Blandt de, der har brugt rettegrupperne, viser der sig nogle forskelle på tværs af fag. Fransk censorerne bruger eksempelvis rettegruppen meget, da der overordnet er få fransk censorer, som derfor har et stort behov for at være knyttet til hinanden i forhold til sparring mv. Det gør, at de kender hinanden godt og sparrer om stort og småt. I de øvrige fag er det mere sporadisk, om censoren har deltaget i en rettegruppe. Eksempelvis beskriver en matematikcensor, at han bruger rettegruppen, da hans oplevelse er, at censorkollegerne i rettegruppen giver mere kvalificeret sparring end de andre matematiklærere på skolen, som har mindre erfaring end vedkommende selv.

Det tyder imidlertid på, at sparringen om opgaverne har en vis værdi, idet 97 % af censorerne i et eller andet omfang oplevede, at sparringen hjalp dem, jf. Tabel 4.6. Svarfordelingerne på spørgsmålet om, i hvilket omfang sparringen hjalp, indikerer, at i tre ud af fire tilfælde hjalp sparringen med over halvdelen af besvarelserne. Dette vidner om, at sparringsmuligheden ikke kan stå alene, da hver fjerde havde behov for at søge yderligere sparring.

Tabel 4.6 Censorbesvarelser på spørgsmålet "Hjalp sparringen dig med at blive mere sikker på, hvilken karakter du skulle give?" Antal og procent.

	Antal	Procent
Sparringen hjalp mig i forbindelse med alle de diskuterede besvarelser	82	53,6
Sparringen hjalp mig med 76-99 % af besvarelserne	19	12,4
Sparringen hjalp mig med 51-75 % af besvarelserne	14	9,2
Sparringen hjalp mig med 26-50 % af besvarelserne	2	1,3
Sparringen hjalp mig med 1-25% a af besvarelserne	31	20,3
Jeg sparrede, men det hjalp mig ikke	5	3,3
Antal besvarelser	153	100

Kilde: VIVEs spørgeskema til censorer.

De øvrige censorer nævner forskellige grunde til, at de ikke bruger rettegrupperne. Nogle få er sjældent i tvivl og har derfor sjældent brug for sparring. Langt de fleste censorer på tværs af fag er dog enige om, at det er nødvendigt med sparring, men de sparrer i stedet med lærerkolleger på deres skole eller censorer i deres netværk, som de fx kender fra censorkurser. Rettegrupperne ses altså som et ringere alternativ til den sparring, de ellers benytter sig af. Nogle mener, at rettegrupper kan være en god ide for nye censorer eller censorer uden andre sparringsmuligheder.

Interviewene viser, at det opleves som en væsentlig barriere, at censorerne ikke kender de andre i rettegruppen. De vil gerne undgå at "ulejlige" en fremmed, som har travlt med at rette sine egne opgaver, men vil hellere spørge nogle, de kender og har naturlig kontakt med. De censorer, som selv er blevet kontaktet af nogle fra rettegruppen, følte sig dog ikke ulejliget. Et andet eksempel er en dansk censor, der finder det uoverskueligt at skulle forklare sine tanker om en opgave til en fremmed, som måske har et helt andet vurderingsgrundlag. Praktikken med at scanne og sende opgaver beskrives desuden som "bøvlet" og "uoverskuelig" i forhold til at "tage to bedømmelser under armen" og tale med en kollega, man alligevel ser. Det nævnes også, at sparringen med rettegruppen er ufleksibel:

Det dur ikke til mit temperament at skulle vente 14 dage på at snakke med nogen om en opgave, fordi der er en bestemt kommunikationsdag. Jeg har brug for at sparre med det samme. (Censor i dansk)

For langt de fleste censorer foregik sparringen enten telefonisk (49 % i dansk, 45 % i matematik og 39 % i sprogfagene) eller via e-mail (36 % i dansk, 34 % i matematik og 56 % i sprogfagene), jf. Bilagstabel 2.9. 56 % af censorerne har sendt de relevante besvarelser til hinanden over mail, mens 30 % valgte slet ikke at sende besvarelserne (jf. Bilagstabel 2.8).

Blandt skolelederne er der overvejende opbakning til, at skolens lærere, der agerer som censorer, får den arbejdstid, de bruger på at deltage i rettegrupper enten dækket fuldstændig eller for det meste, jf. Tabel 4.7. Dette svarer 54 % af skolelederne på folkeskoler, mens 63 % af skolelederne på frie og private grundskoler svarer dette. Andelen af skoleledere på efterskoler, der bevilger fuldstændig eller næsten fuldstændig arbejdstid er på 42 % Tilsvarende svarer hele 63 % af skolelederne, at de opfordrer censorerne til at deltage i rettegrupperne, og at 9 % stiller krav om, at de deltager i rettegrupper (jf. Bilagstabel 2.11).

Tabel 4.7 Skolelederbesvarelser på spørgsmålet "Har de lærere på skolen, der er beskikkede censorer i de skriftlige prøver i 9. eller 10. klasse, arbejdstid til at deltage i rettegrupper?" Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler	Total
Fuldstændig dækning af arbejdstid	34,2	40,7	26,9	34,1
Det meste dækket	20,3	22,2	15,4	19,7
En mindre del dækket	12,7	3,7	15,4	11,4
Slet ikke	7,6	25,9	23,1	14,4
Ingen censorer har benyttet sig af dette tilbud	25,3	7,4	19,2	20,5
Antal besvarelse	79	27	26	132

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,129. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til skoleledere.

Behovet for sparring varierer på tværs af de fag, censorerne er tilknyttet. I matematik knytter sparring sig typisk til det overordnede eller en delopgave frem for en hel besvarelse. I dansk indeholder bedømmelsen et større element af skøn, hvorfor sparringen især bruges til at få andre perspektiver på en bestemt besvarelse. I sprogfagene har censorerne behov for sparring om generelle tvivl eller formelle ting, fx hvad man gør med billeder i opgaver, snarere end den enkelte besvarelse. I tysk og fransk hjælpes man desuden ad med at opdage og håndtere snyd.

4.5.1 SkoleKom

En alternativ sparringsmetode, som er relativt udbredt blandt censorer, består i at stille specifikke, opgaverrelaterede spørgsmål på lærernes videndelingsplatform, SkoleKom, hvor censorerne har adgang til specifikke censorfora, de såkaldte "censorkonferencer". Tabel 4.8 viser, at der er stor forskel på, i hvilken grad censorer i de forskellige fag gør brug af censorkonferencen.

Tabel 4.8 Censorbesvarelser på spørgsmålet: "I hvilken grad bruger du censorkonferencen på SkoleKom som sparringsforum?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
I høj grad	36,8	71,0	36,0	49,9
I nogen grad	35,7	21,3	32,3	29,3
I mindre grad	21,9	7,3	18,8	15,5
Slet ikke	5,7	0,3	12,9	5,3
Antal besvarelser	283	300	186	769

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,000. Dermed er der statistisk signifikante forskelle på svarfordelingerne
 Kilde: VIVEs spørgeskema til censorer.

Særligt matematikcensorerne gør i nogen eller høj grad brug af konferencen, og de kvalitative interview viser også, at netop denne censorgruppe er glade for SkoleKom:

På den konkrete opgave er det bedre at spørge ud på konferencen [på SkoleKom, red.], hvor der lynhurtigt kommer svar, frem for at skulle ulejlige en bestemt person, som måske ikke har tid, og hvor man kun får én persons mening. (Censor i matematik)

Matematikcensorerne bruger platformen til at debattere og afklare tvivlsspørgsmål i forbindelse med bedømmelsesperioden. Her fortæller de, at man kan få mange svar hurtigt fra kolleger eller læringskonsulenter. De er dog enige i, at det kræver, at der er en dedikeret person, der er ansvarlig for at rydde op i samtalerne, da det let kan blive rodet eller ende i 'skænderier'. I modsætning hertil beskriver flere danskensorer, at de ikke begejstrede for SkoleKom som platform, netop fordi de har oplevet, at der er for rodet og usystematisk og mest bruges til irrelevante snakke. Dette afspejler sig i høj grad også i censorernes besvarelser på, i hvilken grad de lægger vægt på sparringen fra censorkonferencen i deres bedømmelser, jf. Tabel 4.9.

Tabel 4.9 Censorbesvarelser på spørgsmålet: "I hvilken grad lægger du vægt på sparringen fra censorkonferencen på SkoleKom i dine bedømmelser?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
I høj grad	30,1	60,3	29,0	41,7
I nogen grad	40,8	30,3	33,9	35,0
I mindre grad	24,1	8,0	21,0	17,1
Slet ikke	5,0	1,3	16,1	6,3
Antal besvarelser	282	300	186	768

Anm.: Svarfordelingerne er testet med Chi²-test: p=0,000. Dermed er der statistisk signifikante forskelle på svarfordelingerne.
Kilde: VIVEs spørgeskema til censorer.

Det overordnede billede fra det kvalitative interviewmateriale er imidlertid, at SkoleKom ikke anvendes systematisk i forbindelse med rollen som censor.

4.6 Censorkurser og webinarer

Undervisningsministeriet beskikker censorer efter indstilling fra skolens leder, der er ansvarlig for at påse, at de censorer, der indstilles, er kvalificerede. De indstillede censorer skal have gennemført et fuldt undervisningsforløb, der fører frem til den pågældende prøve, inden for de sidste tre år. Derudover skal man mindst hvert tredje år have deltaget i de censorkurser, som Styrelsen for Undervisning og Kvalitet (STUK) afholder. Mens deltagelse i censorkurserne er en forudsætning for beskikkelse, kan de også løbende anvendes til at repetere vigtig viden for censorer, der allerede er blevet beskikkede. Ud over censorkurserne afholder STUK også en række webinarer (online seminarer). I maj 2018 blev der som et nyt tiltag afholdt webinarer efter prøverne for censorer i fagene dansk, matematik og tysk. Deltagelse var dog ikke var obligatorisk. I dette delkapitel undersøges omfanget af censorernes deltagelse i censorkurser såvel som disse webinarer.

Hvad angår censorkurserne er der overordnet set meget god tilslutning. Tabel 4.10 viser, at 94-95 % af censorerne i matematik og dansk deltog i et censorkursus i løbet af skoleåret 2017/2018. Dette gjaldt for 89 % af censorerne i sprogfagene.

Tabel 4.10 Censorbesvarelser på spørgsmålet: "Har du i løbet af skoleåret 2017/2018 deltaget i et af Styrelsen for Undervisning og Kvalitets censorkurser i det fag, du var censor i?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
Ja	95,0	94,0	89,3	93,2
Nej	5,0	6,0	10,7	6,8
Antal besvarelser	282	300	187	769

Anm.: Svarfordelingerne er testet med Chi²: p=0,043. Dermed er der statistisk signifikante forskelle på svarfordelingerne.
Kilde: VIVEs spørgeskema til censorer.

De fleste censorer fandt censorkurserne brugbare i forhold til deres bedømmelser, hvor 75 % af danskcensorerne, 73 % af matematikcensorerne og 73 % af sprogfagscensorerne i høj grad fandt censorkurset brugbart, jf. Bilagstabel 2.12. I gennemsnit fandt blot 4 % af censorerne i mindre grad kurset brugbart. Her oplevede flest sprogfagslærere, at det kun i mindre grad var brugbart. Deltagelse i censorkurser er noget, de fleste skoleledere vil bevilge enten fuld arbejdstid til, eller det meste af tiden dækket (mellem 78 og 89 % af skolelederne, jf. Bilagstabel 2.13).

De kvalitative interview understøtter dette, da censorerne i interviewene overvejende er positive over for både censorkurser og webinarer. Censorerne bruger censorkurserne og webinarerne som forberedelse til den kommende prøve og som bistand i forbindelse med retning af opgaver, men i høj grad også som en måde at følge med i, hvor deres fag bevæger sig hen og for at få input til egen undervisning.

Vi er jo censorer for at følge med. (Censor i matematik)

Det er tydeligt på tværs af fag, at censorerne i høj grad oplever, at censorkurset bidrager til forståelse for opgaven som censor. Flere censorer udtrykker dog, at der er plads til forbedring. Eksempelvis er der et ønske om, at censorkurset skal ligge tidligt nok til, at de kan indgå bedre i planlægningen af skoleåret, så informationen kan nå ud til lærerne tidligere. På kurserne indgår ofte viden, som også er relevant for lærerne i forhold til deres forberedelse af undervisningen. Eksempelvis forklarer en censor i matematik, at man har oplevet at finde ud af i januar, at man ikke længere måtte bruge internettet i forbindelse med den skriftlige prøve. Og det betyder, at lærerne i det første halve år potentielt har arbejdet uhensigtsmæssigt i forhold til inddragelse af internettet i problemløsningen.

Tilsvarende er der overvejende positive holdninger til indholdet på dagen for censorkurset, hvor et mindretal oplever, at der i anden halvdel af dagen fokuseres for meget på dialog og snak, og at man derfor med fordel kunne gøre dagen kortere. Særligt matematikcensorerne oplever, at formiddagen på censorkurset er mere relevant end eftermiddagen. Programmet for eftermiddagen er der delte meninger om – nogle er glade for muligheden for at drøfte og regne konkrete eksempler med andre, mens andre føler, at dagen bliver meget lang, og at eftermiddagen i højere grad bidrager med kendt stof. Det fysiske møde giver bedre mulighed for at drøfte problemstillinger i en dialog med andre og giver bedre perspektiver på indholdet.

Jeg synes, at vægtningen [af dagen, red.] bør være anderledes. Prøverettedelen bør vægtes mere. Jeg tager [stilene, red.] med hjem til mine kollegaer, hvor vi sidder i fællesskab og retter og sammenligner deres bedømmelser med den givne. Det er en del af mit hverv som censor at opdatere kollegaer, som ikke er lige så heldige at være censorer. (Censor i fransk)

Censorkurset giver også indsigt i ministeriets tanker om prøven og læringskonsulenternes vejledning til, hvad censorerne skal kigge på.

Læringskonsulenterne i dansk er knivskarpe og har en fast plan for, hvad de skal sige, og hvad man skal være opmærksom på, det bruger jeg meget. Og de tager udgangspunkt i opgaveformuleringerne fra året før og taler om, hvordan det er blevet tolket, om det er blevet forstået rigtigt, og om opgavekommissionen har formuleret dem på en ordentlig måde. (Censor i dansk)

Der er dog langt overvejende en oplevelse af, at kurset er vigtigt, og at det er vigtigt at mødes fysisk for at sikre den bedst mulige overlevering af information samt drøftelsen af disse.

4.6.1 Webinarer

Hvad angår STUKs webinarer, afholdes disse på udvalgte tidspunkter. Deltagelse i disse er ikke obligatorisk, men kan ligesom censorkonferencerne være med til at opklare eventuelle tvivlsspørgsmål i forbindelse med bedømmelserne. Webinarerne faciliteres af fx en læringskonsulent fra STUK og kan streames enten live, eller efter webinarer er afsluttet. Webinarerne er mindre anvendte blandt censorerne end censorkurset, jf. Tabel 4.11. Sammenligner man på tværs af fag er det i videst udstrækning danskensorer, der deltager i webinarerne.

Tabel 4.11 Censorbesvarelser på spørgsmålet: "Efter de skriftlige prøver i maj 2018 afholdt Styrelsen for Undervisning og Kvalitet webinarer for skriftlige censorer i nogle fag. Deltog du i et eller flere af disse webinarer for skriftlige censorer i dit fag?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
Ja, jeg så det live	36,2	20,0	25,9	27,7
Ja, men jeg så det ikke live	27,0	18,7	5,2	21,1
Nej	36,9	61,3	69,0	51,3
Antal besvarelser	282	300	58	640

Anm.: $p=0,000$

Anm.: Svarfordelingerne er testet med Chi2: $p=0,000$. Dermed er der statistisk signifikante forskelle på svarfordelingerne. Der blev ikke afholdt webinar for engelsk og fransk.

Kilde: VIVEs spørgeskema til censorer.

Den noget lavere deltagelse i webinarerne sammenlignet med censorkurserne skyldes ifølge interviewpersonerne forskellige årsager, eksempelvis at det ligger i en periode, hvor der er meget travlt, at der var it-mæssige udfordringer i forhold til at koble sig på webinarer, og at censorerne følte, at de havde godt styr på materialet i forvejen. Der er dog umiddelbart bred tilslutning blandt de censorer, der har benyttet sig af webinarer, at det var værdifuldt, brugbart og gav mening.

Det gode ved webinarer er, at de slavisk gennemgår rettevejledningen. Det er meget konkret, og det giver rigtig meget. Det giver særligt noget i forhold til, hvis man er i tvivl. ... [Læringskonsulenten, red.] pinder alt ud i hver af de opgaveoplæg, som ligger der for både FP9 og FP10. Det gør, at vi som censorer er endnu mere klædt på. [Læringskonsulenten, red.] har selv siddet og rettet opgaverne, så hun ved, hvad det handler om. Hun er virkelig klædt godt på. (Censor i dansk)

Flere efterspørger, at webinarer ligger tidligere, end det gør på nuværende tidspunkt. Det er dog en balancegang, da der på den ene side er behov for, at man som censor er godt inde i opgaverne inden webinarer, men ikke for langt igennem bunken, da man så oplever, at hjælpen og vejledningen kommer for sent.

Jeg kunne godt tænke mig, at webinarer lå tidligere. Jeg var igennem i hvert fald halvdelen af rettelserne, før det blev afholdt. (Censor i dansk)

En censor i matematik beskriver, at fordi mange fravælger webinarer, bør man sikre, at censorkurset er det bærende element. Samtidig udtrykker samme censor også, at det kan være problematisk, at mange tænker, at de "ikke har brug for webinarer, fordi de har styr på det".

4.7 Delkonklusion

Både censorer og skoleledere ser det som en fordel for de enkelte lærere og for skolerne at være/have beskikkede censorer. Hvervet ses som en god mulighed for faglig opdatering og udvikling, og censorerne ses generelt som gode sparrings- og informationskilder på skolen. Alligevel er det skoleledernes vurdering, at der ikke er ret stor interesse for hvervet blandt lærerne. På tværs af både censorer, skoleledere og skolelærere er opfattelsen, at mange lærere vælger censorhvervet fra, fordi det er meget arbejde på et tidspunkt af året, hvor man i forvejen har travlt. Der er desuden flere lærere, som ikke finder opgaven ret interessant. Det blev i forbindelse med et interview foreslået, at man kunne afholde "førcensorkurser", hvor man kunne udbrede fordelene ved hvervet for at øge interessen. Desuden virker det vigtigt at understrege muligheden for at rette en mindre mængde af opgaver, samt at skolelederne udtrykker opbakning til hvervet. For mange af skolelederne er fordelene tydelige, men det er ikke nødvendigvis ensbetydende med, at de oplever et ansvar for at motivere og kvalificere censorerne til hvervet. Fordi censorerne er beskikket og aflønnet af ministeriet, ses kvalificeringen af arbejdet også som ministeriets opgave. Hvis opgaven skal ligge på skolen, er den overordnede holdning, at opgaven skal understøttes af ressourcer.

Samarbejdet mellem censorer og skolerne er begrænset men fungerer overvejende godt. Langt de fleste interaktioner om praktik eller klagesager foregår gnidningsfrit og efter relativt faste procedurer. Der efterspørges blandt såvel censorer som skoleledere en mere fast procedure for det praktiske samarbejde med hensyn til fremsendelse af opgaver og kontaktoplysninger mv., da dette vil lette arbejdet samt gøre det nemmere at huske fra gang til gang. I forhold til klagesager er de gældende procedurer velfungerende, om end de i visse tilfælde føles lidt langsommelige. Inden for både det praktiske samarbejde og samarbejdet vedrørende klagesager er der dog også eksempler på dårligt samarbejde, hvor typisk den ene part oplever, at den anden part ikke håndterer opgaven tilfredsstillende. Dette kan skyldes menneskelige fejl i bedømmelser eller personbårne samarbejdsvanskeligheder.

Censorerne fra de forskellige fag anvender de tilgængelige bedømmelsesværktøjer forskelligt og tillægger dem forskellig værdi. Analysen viser, at dansk censorerne i særlig grad anvender vurderingsskemaet, der tager udgangspunkt i den vejledende karakterbeskrivelse, da det er med til at give en ensartet bedømmelsespraksis. Matematikcensorer anvender i særlig grad rettevejledningen. Blandt censorerne i sprogfagene anvendes primært vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen. Fælles er dog, dels at de generelt vurderer, at bedømmelsesværktøjerne er brugbare, dels at de bidrager til større gennemsigtighed og objektivitet i forbindelse med bedømmelserne, hvilket også lægges til grund, når de vælger et konkret værktøj. Censorerne giver udtryk for, at det er et solidt fagligt grundlag at foretage deres bedømmelse på.

Rettegrupperne bliver kun brugt i mindre grad af censorerne. Cirka hver femte censor har en eller flere gange gjort brug af rettegrupperne. Dette skyldes ikke ringe kvalitet i rettegrupperne, da de bliver vurderet til at hjælpe, men i højere grad, at censorerne bruger andre former for sparring, særligt blandt kolleger, som de i forvejen kender og har samarbejdet med før eller eksempelvis censor-konferencen på SkoleKom. Det er en barriere, at censorerne ikke kender hinanden i rettegrupperne. Ligeledes udtrykker flere, at de ikke oplever et behov for at bruge rettegrupperne til sparring.

Analysen viser, at STUKs censorkurser både bliver brugt, og vurderes til at være gode til at kvalificere censorernes arbejde. Censorkurset bruges til at holde sig opdateret på fagets udvikling og den kommende prøve samt på at få input til sin egen undervisning. Flere udtrykker et ønske om, at kurset ligger tidligere på skoleåret, så indholdet i højere grad ville kunne overføres til skolerne og lærerne, og dermed i højere grad indgå i undervisningen op til prøven. Derudover viser analysen, at STUKs

webinarer bliver brugt i mindre grad, men at de også vurderes til at være gode. De bruges mindre, da de ikke er obligatoriske, og da mange oplever, at de har meget travlt på det tidspunkt, hvor webinarerne gennemføres. Dertil kommer, at en del censorer ikke oplever behov for at deltage i webinarerne.

Overordnet viser analysen af de forskellige bedømmelsesværktøjer og fora, at censorerne har mange muligheder for at forberede sig til prøven samt få sparring på eventuelle spørgsmål. Der udtrykkes ikke en efterspørgsel efter nye eller andre former for sparringsmuligheder eller værktøjer.

Bilag 1 Metode

Evalueringen af ordningen med én bedømmer benytter sig af både kvantitative og kvalitative metoder. Den kvantitative del af undersøgelsen består af spørgeskemaer sendt til censorer og skoleledere samt en statistisk analyse af ordningen med én bedømmers eventuelle indflydelse på karakterudviklingen. Den kvalitative del af undersøgelsen består af en række enkelt- og fokusgruppeinterview med skoleledere, censorer og lærere. I dette afsnit beskriver vi metoden og datagrundlaget for de to dele af undersøgelsen særskilt.

Kvantitativ del af undersøgelsen

Surveys

Den kvantitative del af undersøgelsen består af to separate spørgeskemaundersøgelser, til henholdsvis censorer og skoleledere. Undersøgelsen blandt censorer er udført som en totalundersøgelse blandt alle beskikkede censorer, dvs. spørgeskemaet er sendt ud til 1.155 censorer. Heraf svarede 773 censorer på skemaet. Fordelingen af besvarelser fremgår af Bilagstabel 1.1.

Bilagstabel 1.1 Svarprocenter for censorsurvey. Antal og procent.

	Stikprøve	Hele besvarelser	Hele og delvise besvarelser	Svarprocent	Svarprocent inkl. delvise
Dansk	442	285	300	64,5	67,9
Matematik	433	300	325	69,3	75,1
Engelsk	181	123	129	68,0	71,3
Fransk	10	7	9	70,0	90,0
Tysk	89	58	61	65,2	68,5
I alt	1.155	773	824	66,9	71,3

Kilde: VIVEs spørgeskema til censorer.

Sammen med spørgeskemaerne har censorerne modtaget en invitation til at deltage i undersøgelsen på vegne af Styrelsen af Undervisning og Kvalitet (STUK). Der er også blevet gjort opmærksom på undersøgelsen på censorkonferencerne på SkoleKom. Der er i forbindelse med dataindsamlingen udsendt to rykkermails til respondenter, der enten ikke har påbegyndt besvarelserne af spørgeskemaet, eller som kun har svaret delvist.

Foruden censorer, har også skoleledere på folkeskoler, frie og private grundskoler samt efterskoler deltaget i en spørgeskemaundersøgelse. Skolelederne er valgt som repræsentanter for skolernes oplevelser med ordningen. Undersøgelsen blandt efterskoleledere er udført som en totalundersøgelse, mens en simpel tilfældig stikprøve er udtrukket for ledere på henholdsvis folkeskoler og frie og private grundskoler. Stikprøvestørrelserne er beregnet ud fra en forventning om en svarprocent på ca. 50 %.

Bilagstabel 1.2 Svarprocenter for skoleledere. Antal og procent.

	Stikprøve	Hele besvarelser	Hele og delvise besvarelser	Svarprocent	Svarprocent inkl. delvise
Folkeskoler	547	229	237	41,9	43,3
Frie og private grundskoler	242	125	129	51,7	53,3
Efterskoler	201	98	99	48,8	49,3
I alt	990	452	465	45,7	47,0

Kilde: VIVEs spørgeskema til skoleledere.

Ligesom for censorerne har skolelederne sammen med spørgeskemaet modtaget en skriftlig invitation fra STUK samt to efterfølgende rykkermails vedrørende deltagelse. Da spørgeskemaet til skolelederne har haft til formål at give et indtryk af skolens samlede oplevelser med ordningen, har enkelte skoler, der selv har efterspurgt det, fået tilladelse til, at andre end skolelederen har kunnet besvare skemaet. I tilfælde af disse skoler har vi hos nogle skoler forsikret os om, at det så er den prøveansvarlige, fx viceskolelederen, der har besvaret skemaet. Datagrundlaget for spørgeskemaundersøgelserne vurderes til at være stærkt på grund af antallet af respondenter og den relativt høje svarprocent.

Registeranalyse af karakterudvikling

Foruden spørgeskemaundersøgelsen har vi undersøgt, hvilken indflydelse ordningen med én bedømmer har haft på karakterudviklingen. Baggrunden for denne undersøgelse er data fra Undervisningsministeriets karakterregistre (tilgængeligt gennem uddannelsesstatistik).

De statistiske analyser anvender et Difference-in-Differences design (Angrist & Pischke, 2009) og gennemføres som lineære regressioner med skole-fixed-effects. Designet sammenligner ændringer i to udviklinger over tid, her forstået som udviklingen i karaktergivning sammenlignet med udviklingen i standpunktskaraktergivning. Skole-fixed-effects tager højde for de særlige karakteristika, der knytter sig til hver enkelt skole. Det har den fordel, at modellerne tager højde for det, der skaber de specifikke rammer for undervisningen og karaktergivningen på den enkelte skole. Ydermere tager det også højde for den måde, fx folkeskolereformen er blevet implementeret på skolen. Der kan dog ikke på baggrund af analyserne konkluderes entydigt om kausalitet.

Kvalitativ del af evalueringen

Den kvalitative del af evalueringen bestod i en række semi-strukturerede enkelt- og fokusgruppeinterview. Der er afholdt interview med både censorer, skoleledere og skolelærere. Enkeltinterview er foretaget som telefoninterview.

Fokusgruppeinterviewene har den fordel, at interaktionen mellem deltagerne ofte får dem til at aktivere et langt bredere spænd af deres erfaringer, end hvis der udelukkende gennemføres enkeltinterview. Det kan imidlertid have den ulempe, at det kan være sværere at tale om kontroversielle emner i fokusgruppeinterview end i enkeltinterview, pga. fx normbrud, rolleusikkerhed, tvivl osv. Derfor kan enkeltinterview opleves som tryggere. Idet begge interviewmetoder har hver deres fordele og ulemper, opstiller vi et tosporet design, der benytter sig af både fokusgruppeinterview og enkeltinterview.

Følgende Bilagstabel 1.3 giver et indblik i afholdte interview:

Bilagstabel 1.3 Gennemførte interview. Fordelt på respondenttype. Antal.

	Fokusgruppeinterview	Telefoninterview
Censorer	4	7
Skoleledere	2	7
Skolelærere	4	5
I alt	10	19

Dette tosporede design giver mulighed for, at respondenter, der enten ikke har mulighed for eller ikke er villige til at rejse for at deltage i fokusgruppeinterview, men som alligevel ønsker at interviewes, kan indgå i evalueringen. Desuden giver den tosporede strategi os mulighed for at opnå en geografisk spredning af interviewpersoner, så respondenter fra hele landet får mulighed for at deltage i evalueringen.

Tæller vi alle interviewpersoner med (til både fokusgruppeinterview og telefoninterview), har vi afholdt interview med 40 censorer, 18 skoleledere og 28 skolelærere. I alt indgår altså 86 interviewpersoner. I det følgende gennemgår vi udvælgelseskriterierne for deltagere til interview for henholdsvis censorer, skoleledere og skolelærere.

Udvælgelse af censorer til de kvalitative interview

Der er afholdt fire fokusgruppeinterview med censorer samt syv individuelle telefoninterview. Udvalgsstrategien for censorerne til fokusgruppeinterviewene har været følgende:

Geografisk spredning og minimering af transport for interviewdeltagerne

Der blev afholdt to fokusgruppeinterview i Vestdanmark og to fokusgruppeinterview i Østdanmark. I Østdanmark blev ét interview afholdt i Region Hovedstaden og ét interview i Region Sjælland. I Vestdanmark blev der afholdt ét interview i Region Midtjylland og ét interview i Syddanmark⁶. I hver region deltog censorer fra tre til fem forskellige kommuner.

Minimering af transport for interviewdeltagerne

For at minimere transporten for de deltagende censorer blev kommunerne i hver region inddelt i klynger af tætliggende kommuner. Der er herefter valgt tilfældigt mellem disse klynger

Fagspredning

Der er rekrutteret censorer, således at der på tværs af de fire fokusgruppeinterview sikredes deltagelse fra censorer i både dansk, matematik, engelsk, tysk og fransk.

Fokusgruppeinterviewene blev suppleret med syv individuelle telefoninterview. Seks af censorerne blev valgt ud fra et hensyn til fagspredningen. For at sikre erfaringer fra censorer, der har gjort brug af rettegupperne, blev der derudover udvalgt én censor til telefoninterview ud fra en tilkendegivelse i spørgeskemaundersøgelsen om netop at have gjort brug af sådanne. Denne censor er bedømmer i dansk.

⁶ Der er valgt tilfældigt mellem region Midtjylland, region Nordjylland og Syddanmark (Fyn var ikke en del af udvælgelsen). Udvalgsstrategien er sket vha. random.org

Den samlede fordeling af censorer i forhold til fag er således:

Bilagstabel 1.4 Fordelingen af censorer i forhold til fag.

Dansk	Matematik	Engelsk	Tysk	Fransk	I alt
14	12	6	6	2	40

Fagspredningen er anvendt i alle de kvalitative analyser til at belyse, om der er systematiske forskelle mellem censorernes udtalelser i forskellige fag. Det bemærkes dog, at der er tale om relativt få repræsentanter for hvert fag, herunder særligt fransk. Den kvalitative analyses generaliserbarhed er derfor begrænset.

Udvælgelse af skoleledere til de kvalitative interview

Der er afholdt to fokusgruppeinterview og syv individuelle telefoninterview med skoleledere. Udvælgelsesstrategien⁷ for skolelederne til fokusgruppeinterviewene har været følgende parametre:

Geografisk spredning

Der er afholdt ét fokusgruppeinterview i Vestdanmark og ét fokusgruppeinterview i Østdanmark.

Karakterniveau

Der er tilstræbt variation i forhold til inddragelse af ledere af skoler med henholdsvis lave (< 6,1), middel (6,9-7,2) og høje (>8,1) karakterniveauer i skoleåret 2016/2017. Skolerne er udvalgt fra kommuner, som ellers lå på (eller tæt på) landsgennemsnittet i skoleåret 2016/2017⁸.

Skoletype

Der er tilstræbt variation i forhold til inddragelse af ledere fra forskellige skoletyper. I fokusgruppeinterviewet i Vestdanmark indgår skoleledere fra folkeskoler og efterskoler. I fokusgruppeinterviewet i Østdanmark indgår skoleledere fra folkeskoler. I telefoninterviewene er der suppleret med ledere fra frie grundskoler. Inden for hver skoletype er udvalgt skoler, der varierer i forhold til karakterniveauet (som beskrevet for kriteriet ovenfor).

Minimering af transport for interviewdeltagerne

Det er efterstræbt at få skoler med variation på karakterniveau og skoletype med i undersøgelsen, og som samtidig skoler, der ligger så tæt på hinanden som muligt.

Fokusgruppeinterviewene er suppleret med individuelle telefoninterview. For at sikre deltagelse fra de frie grundskoler, er der først og fremmest udvalgt ledere til telefoninterviewene fra denne skoletype. For også at opnå en dybere forståelse for, hvorfor lederne er henholdsvis tilfredse eller utilfredse med den nye bedømmelsesordning, er der udvalgt to skoleledere, som i spørgeskemaundersøgelsen har angivet høj tilfredshed med ordningen (disse ledere er fra henholdsvis en folkeskole og en efterskole), samt to skoleledere, som har angivet høj utilfredshed med ordningen (disse ledere er fra henholdsvis en folkeskole og en efterskole).

⁷ Data, der ligger til grund for udvælgelsen, stammer fra Styrelsen for It & Læring

⁸ Landgennemsnittet for skoleåret 2016/2017 for de inkluderede skoletyper er på 7,1. I den endelige udvælgelse indgår kommuner med karaktergennemsnit fra 7,1-7,3. I hver kommune udvælges skoler, som varierer i forhold til karaktergennemsnittet. Karakterdata baserer sig på skolernes indberetninger til Styrelsen for It og Læring

Den samlede fordeling af skoleledere i forhold til skoletype og karakterniveau samt tilfredshed med ordningen er således:

Bilagstabel 1.5 Deltagende skoleledere, fordelt på skolens karakterniveau og skoletype. Antal.

	Folkeskoler	Efterskoler	Frie grundskoler	I alt
Lavt niveau (< 6,1)	4	1	1	6
Middel niveau (6,9-7,2)	2	1		3
Højt niveau (> 7,9)	2	1	2	5
Meget tilfreds med ordningen	1	1	-	2
Meget utilfreds med ordningen	1	1	-	2
I alt	10	5	3	18

Opdelingen på skoletype og karakterniveau er anvendt i de kvalitative analyser om skoleledernes overordnede vurdering af bedømmelsesordningen, deres vurdering af censorrollen samt arbejde med at finde beskikkede censorer. Skoletype og karakterniveau kan fx have en betydning for, hvor fokuserede skolerne er på karakterniveauet. Det bemærkes dog, at der er tale om relativt få repræsentanter for de forskellige udvælgelsesparametre. Den kvalitative analyses generaliserbarhed er derfor begrænset.

Udvælgelse af skolelærere til interview

Der afholdes fire fokusgruppeinterview med skolelærere samt fem telefoninterview. For at sikre deltagelse fra skolelærerne, er de først og fremmest udvalgt ud fra skoler, hvor skolelederne også deltager i fokusgruppeinterviewene. Med andre ord har vi bedt skolelederen stille lærere til rådighed for interviewene. I udvælgelsen af lærere har der været fokus på at sikre deltagelse fra lærere med forskellige fagkombinationer. Fokusgruppeinterviewene er suppleret med lærere fra flere forskellige kommuner og fra frie grundskoler. Derudover er fokusgruppeinterviewene suppleret med telefoninterview, hvor der er også er taget højde for skoletype og fagspredning.

Fordelingen af skolelærere på skoletype er således:

Bilagstabel 1.5 Fordelingen af skolelærere på skoletype

Folkeskoler	Efterskoler	Frie grundskoler	I alt
13	7	8	28

Fordelingen i forhold til skoletype er anvendt i de kvalitative analyser til at undersøge, om skolelærere fra forskellige skoletyper varierer i forhold til deres overordnede vurdering af bedømmelsesordningen. I alle de kvalitative analyser med skolelærere er der derudover lagt vægt på udtalelser og refleksioner over, hvorvidt der er noget særligt gældende for lærere fra forskellige fag.

Den kvalitative analyse baseres på et metodisk bredt udvalg af interessenter og aktører samt et væsentligt antal respondenter samlet. Det gør det metodiske grundlag for den kvalitative analyse stærkt. Der er dog analytiske begrænsninger inden for de enkelte grupper af respondenter, hvor der er tale om ganske få respondenter.

Bilag 2 Bilagstabeller

Bilagstabel 2.1 Skolelederbesvarelser på spørgsmålet: "Hvilke grunde oplever du til, at lærerne på din skole kun i mindre grad eller slet ikke er interesserede i at blive beskikede som censorer i skriftlige prøver i 9. eller 10. klasse? (Det er muligt at sætte flere kryds)" Antal og procent.

	Antal	Procent
Det er besværligt at være censor	49	9,7
Det er ikke attraktivt nok at være censor rent fagligt	31	6,2
Betalingen er ikke tilstrækkelig	71	14,1
Der er mange konflikter knyttet til det at være censor	10	2,0
Det er for tidskrævende at være censor	260	51,7
Andre grunde	35	7,0
Ved ikke	47	9,3
Antal observationer	503	100

Anm.: Det har været muligt at sætte flere kryds

Kilde: VIVEs spørgeskema til skoleledere.

Bilagstabel 2.2 Censorbesvarelser på spørgsmålene om den generelle karakterbeskrivelse af 7-trinsskalaen. Fordelt på fag. Procent.

Den generelle karakterbeskrivelse af 7-trinsskalaen...				
	Dansk	Matematik	Sprogfag	Total
<i>... giver mig solidt udgangspunkt for bedømmelsen.</i>				
Enig	79,8	68,0	71,9	73,4
Hverken enig eller uenig	15,7	27,9	20,8	21,6
Uenig	4,5	4,1	7,3	5,1
Antal besvarelser	287	294	192	773
Chi2: p=0,004				
<i>... kunne med fordel uddybes.</i>				
Enig	51,6	43,2	50,3	48,1
Hverken enig eller uenig	34,6	40,1	31,8	36,0
Uenig	13,8	16,7	18,0	16,0
Antal besvarelser	275	287	189	751
Chi2:p=0,188				
<i>... kunne med fordel forenkles.</i>				
Enig	11,5	15,5	18,1	14,7
Hverken enig eller uenig	35,6	47,2	30,2	38,7
Uenig	53,0	37,3	51,7	46,6
Antal besvarelser	270	284	182	736
Chi2:p=0,000				
<i>... er med til at sikre en ensartet bedømmelsespraksis.</i>				
Enig	71,0	60,8	69,7	66,7
Hverken enig eller uenig	21,2	27,0	22,9	23,8
Uenig	7,9	12,3	7,5	9,5
Antal besvarelser	279	293	188	760
Chi2: p=0,069				
<i>... sikrer sammenhæng mellem fælles mål/faglige mål/ målbeskrivelser og min bedømmelse.</i>				
Enig	59,0	53,5	55,6	56,1
Hverken enig eller uenig	33,6	36,7	34,2	35,0
Uenig	7,4	9,8	10,2	9,0
Antal besvarelser	271	286	187	744
Chi2: p=0,065				

Anm.: Svarfordelingerne er testet med Chi2-test.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.3 Censorbesvarelser på spørgsmålene om vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen Fordelt på fag. Procent.

Vurderingskriterierne og den vejledende karakterbeskrivelse i prøvevejledningen...				
	Dansk	Matematik	Sprogfag	Total
<i>... giver mig solidt udgangspunkt for bedømmelsen.</i>				
Enig	94,1	82,8	90,6	88,9
Hverken enig eller uenig	5,6	16,9	6,8	10,2
Uenig	0,4	0,3	2,6	0,9
Antal besvarelser	286	290	192	768
Chi2: p=0,000				
<i>... kunne med fordel udbydes.</i>				
Enig	29,5	26,8	37,6	30,5
Hverken enig eller uenig	45,7	50,7	36,6	45,3
Uenig	24,8	22,5	25,8	24,2
Antal besvarelser	278	280	186	744
Chi2: p=0,040				
<i>... kunne med fordel forenkles.</i>				
Enig	14,4	17,3	15,1	15,7
Hverken enig eller uenig	43,2	52,8	37,6	45,5
Uenig	42,4	29,9	47,3	38,9
Antal besvarelser	271	284	186	741
Chi2: p=0,002				
<i>... er med til at sikre en ensartet bedømmelsespraksis.</i>				
Enig	86,0	72,2	80,4	79,4
Hverken enig eller uenig	10,8	22,6	15,8	16,5
Uenig	3,2	5,2	3,8	4,1
Antal besvarelser	279	288	184	751
Chi2: p=0,002				
<i>... sikrer sammenhæng mellem fælles mål/faglige mål/ målbeskrivelser og min bedømmelse.</i>				
Enig	73,3	64,5	72,8	69,8
Hverken enig eller uenig	23,7	33,0	22,8	27,0
Uenig	3,0	2,5	4,4	3,2
Antal besvarelser	270	279	180	729
Chi2: p=0,058				

Anm.: Svarfordelingerne er testet med Chi2-test.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.4 Censorbesvarelser på spørgsmålene om vurderingsskemaer, der tager udgangspunkt i den vejledende karakterbeskrivelse. Fordelt på fag. Procent.

Vurderingsskema, der tager udgangspunkt i den vejledende karakterbeskrivelse...				
	Dansk	Matematik	Sprogfag	Total
<i>... giver mig et solidt udgangspunkt for bedømmelsen.</i>				
Enig	90,9	76,9	90,0	85,7
Hverken enig eller uenig	7,3	20,1	7,9	12,0
Uenig	1,8	3,0	2,1	2,3
Antal besvarelser	286	264	189	739
Chi2: p=0,000				
<i>... kunne med fordel uddybes.</i>				
Enig	35,3	17,3	36,8	29,2
Hverken enig eller uenig	41,8	59,2	40,7	47,8
Uenig	22,9	23,5	22,5	23,0
Antal besvarelser	275	260	182	717
Chi2: p=0,000				
<i>... kunne med fordel forenkles.</i>				
Enig	11,0	13,0	16,2	13,0
Hverken enig eller uenig	39,9	59,9	38,0	46,8
Uenig	49,1	27,1	45,8	40,2
Antal besvarelser	273	262	179	714
Chi2: p=0,000				
<i>... er med til at sikre en ensartet bedømmelsespraksis.</i>				
Enig	83,6	67,2	75,9	75,7
Hverken enig eller uenig	13,2	28,2	19,8	20,3
Uenig	3,2	4,6	4,3	4,0
Antal besvarelser	280	262	187	729
Chi2: p=0,000				
<i>... sikrer sammenhæng mellem fælles mål/faglige mål/målbeskrivelser og min bedømmelse.</i>				
Enig	70,2	59,8	66,5	65,5
Hverken enig eller uenig	26,8	37,9	29,1	31,5
Uenig	2,9	2,3	4,4	3,1
Antal besvarelser	272	261	182	715
Chi2: p=0,051				

Anm.: Svarfordelingerne er testet med Chi2-test.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.5 Censorbesvarelser på spørgsmålene om rettevejledning for faget. Fordelt på fag. Procent.

Rettevejledning for faget...				
	Dansk	Matematik	Sprogfag	Total
<i>...giver mig et solidt udgangspunkt for bedømmelsen.</i>				
Enig	84,8	98,0	82,3	89,4
Hverken enig eller uenig	8,9	1,7	16,0	7,8
Uenig	6,3	0,3	1,7	2,8
Antal besvarelser	270	296	181	747
Chi2: p=0,000				
<i>... kunne med fordel uddybes.</i>				
Enig	32,1	19,5	22,6	24,8
Hverken enig eller uenig	44,2	35,5	49,2	42,0
Uenig	23,8	45,0	28,3	33,2
Antal besvarelser	265	287	177	729
Chi2: p=0,000				
<i>...kunne med fordel forenkles.</i>				
Enig	16,4	9,4	13,2	12,9
Hverken enig eller uenig	41,6	29,9	48,3	38,5
Uenig	42,0	60,8	38,5	48,6
Antal besvarelser	262	288	174	724
Chi2: p=0,000				
<i>...er med til at sikre en ensartet bedømmelsespraksis.</i>				
Enig	74,2	93,9	74,4	82,1
Hverken enig eller uenig	19,1	4,4	23,3	14,2
Uenig	6,7	1,7	2,3	3,7
Antal besvarelser	267	295	176	738
Chi2: p=0,000				
<i>...sikrer sammenhæng mellem fælles mål/faglige, mål/målbeskrivelser og min bedømmelse.</i>				
Enig	66,3	76,7	66,5	70,5
Hverken enig eller uenig	28,0	20,5	31,2	25,8
Uenig	5,7	2,8	2,3	3,7
Antal besvarelser	264	288	173	725
Chi2: p=0,013				

Anm.: Svarfordelingerne er testet med Chi2-test.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.6 Censorbesvarelser på spørgsmålet: "Som censor har du været en del af en rettegruppe. Havde din rettegruppe kontakt *forud* for rettearbejdet?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
Ja, jeg havde kontakt til én anden fra rettegruppen	13,4	9,0	7,5	10,3
Ja, jeg havde kontakt med flere fra rettegruppen	2,5	4,7	7,5	4,6
Nej, jeg havde ikke kontakt med nogen fra rettegruppen	84,1	86,3	85,0	85,2
Antal besvarelser	283	300	186	769

Anm.: Svarfordelingerne er testet med Chi2-test: $p=0,027$. Dermed er der statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.7 Censorbesvarelser på spørgsmålet: "Hvorfor benyttede du dig ikke af rettegruppen?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
Der var ikke et behov for sparring	42,1	44,4	57,0	46,6
Jeg kendte ikke de andre censorer godt nok	2,9	0,00	0,00	1,0
Jeg ville ikke forstyrre de andre censorer	1,9	0,00	1,4	1,0
Jeg brugte andre sparringsmuligheder	41,2	52,8	36,6	44,9
Andre årsager	12,0	2,8	4,9	6,5
Antal besvarelser	209	252	142	603

Anm.: Spørgsmålet er kun stillet til censorer, der har angivet, at de ikke har haft kontakt med nogen fra rettegruppen. Svarfordelingerne er testet med Chi2-test: $p=0,000$. Dermed er der statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.8 Censorbesvarelser på spørgsmålet: "Sendte du de relevante besvarelser til din sparringspartner, inden I skulle sparre?" Antal og procent.

	Antal	Procent
Ja, jeg sendte dem pr. mail som scannede filer	86	55,8
Ja, jeg lagde besvarelser som scannede filer i en delt folder på nettet (Drop Box, iCloud, etc.)	2	1,3
Ja, jeg tog kopier og sendte dem pr. brev	1	0,7
Ja, jeg brugte en anden metode	19	12,3
Nej, jeg sendte ikke besvarelser	46	29,9
Total	154	100

Anm.: Spørgsmålet er kun stillet til censorer, der har angivet, at de har haft kontakt med nogen fra rettegruppen.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.9 Censorbesvarelser på spørgsmålet: "Hvordan foregik sparringen?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
Vi mødtes fysisk	11,6	6,8	0,0	7,1
Telefonisk	49,3	45,5	39,0	45,5
Virtuelt, fx video-møde på Skype	1,5	9,1	0,0	3,3
E-mail	36,2	34,1	56,1	40,9
Andet	1,5	4,6	4,9	3,3
Antal besvarelser	69	44	41	154

Anm.: Spørgsmålet er kun stillet til censorer, der har angivet, at de har haft kontakt med nogen fra rettegruppen. Svarfordelingerne er testet med Chi2-test: $p=0,035$. Dermed er der statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.10 Skolelederbesvarelser på spørgsmålet: "Har de lærere på skolen, der er beskikkede censorer i de skriftlige prøver i 9. eller 10. klasse, arbejdstid til at deltage i følgende rettegrupper?" Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler	Total
Fuldstændig dækning af arbejdstid	34,2	40,7	26,9	34,1
Det meste dækket	20,3	22,2	15,4	19,7
En mindre del dækket	12,7	3,7	15,4	11,4
Slet ikke	7,6	25,9	23,1	14,4
Ingen censorer har benyttet sig af dette tilbud	25,3	7,4	19,2	20,5
Antal besvarelser	79	27	26	132

Anm.: Spørgsmålet er kun stillet til skoleledere, der har angivet, at de har beskikkede censorer i lærerstaben. Svarfordelingen er testet med Chi2: $p=0,129$. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til skoleledere.

Bilagstabel 2.11 Skolelederbesvarelser på spørgsmålet: "Stiller I krav til eller opfordrer I de lærere, der er beskikkede som censorer, til at deltage i rettegrupper?" Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler	Total
Nej	24,8	31,7	34,2	28,3
Vi opfordrer dem til at deltage	65,7	56,1	60,5	62,5
Vi stiller krav om, at de deltager	9,5	12,2	5,3	9,2
Antal besvarelser	105	41	38	184

Anm.: Spørgsmålet er kun stillet til skoleledere, der har angivet, at de har beskikkede censorer i lærerstaben. Svarfordelingen er testet med Chi2: $p=0,624$. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til skoleledere.

Bilagstabel 2.12 Censorbesvarelser på spørgsmålet: "I hvilken grad fandt du censorkurset brugbart i forhold til dine bedømmelser?" Fordelt på fag. Procent.

	Dansk	Matematik	Sprogfag	Total
I høj grad	75,4	73,0	76,1	74,6
I nogen grad	22,0	22,1	16,8	20,8
I mindre grad	2,6	4,3	6,6	4,2
Slet ikke	0,0	0,7	0,6	0,4
Antal besvarelser	268	281	167	716

Anm.: Spørgsmålet er kun stillet til censorer, der har angivet, at de har deltaget i censorkurset. Svarfordelingerne er testet med Chi2-test: $p=0,269$. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.13 Skolelederbesvarelser på spørgsmålet: "Har de lærere på skolen, der er beskikkede censorer i de skriftlige prøver i 9. eller 10. klasse, arbejdstid til at deltage i følgende aktiviteter? Censorkurser." Fordelt på skoletype. Procent.

	Folkeskoler	Frie grundskoler	Efterskoler	Total
Fuldstændig dækning af arbejdstid	61,1	72,2	47,2	60,5
Det meste dækket	24,2	16,7	30,6	23,0
En mindre del dækket	6,3	5,6	8,3	6,6
Slet ikke	5,3	5,6	11,1	6,6
Ingen censorer har benyttet sig af dette tilbud	3,2	0	2,8	2,4
Antal besvarelser	95	36	36	167

Anm.: Spørgsmålet er kun stillet til skoleledere, der har angivet, at de har beskikkede censorer i lærerstaben. Svarfordelingen er testet med Chi2: $p=0,635$. Dermed er der ikke statistisk signifikante forskelle på svarfordelingerne.

Kilde: VIVEs spørgeskema til skoleledere.

Bilagstabel 2.14 Censorbesvarelser på spørgsmålet: "Hvor enig eller uenig er du i følgende spørgsmål? Jeg ved, hvordan jeg skal agere, hvis en skole klager over en bedømmelse" Antal og procent.

	Antal	Procent
Meget enig	311	40,6
Enig	370	48,2
Hverken enig eller uenig	47	6,1
Uenig	36	4,7
Meget uenig	3	0,4
Antal besvarelser	767	100

Kilde: VIVEs spørgeskema til censorer.

Bilagstabel 2.15 Skolelederbesvarelser på spørgsmålet: "I hvilken grad vurderer du, at dialogen med statsligt beskikkede censorer i forbindelse med klager fungerer tilfredsstillende i forhold til censorers viden om klageproceduren (herunder tidsfrister)?" Antal og procent.

	Antal	Procent
I høj grad	99	47,1
I nogen grad	86	41,0
I mindre grad	19	9,1
Slet ikke	6	2,8
Total	210	100

Kilde: VIVEs spørgeskema til skoleledere.

Bilagstabel 2.16 Skolelederbesvarelser på spørgsmålet: "I hvilken grad vurderer du, at dialogen med statsligt beskikkede censorer i forbindelse med klager fungerer tilfredsstillende i forhold til censorers lydhørhed i forhold til klager?". Antal og procent.

	Antal	Procent
I høj grad	62	29,3
I nogen grad	85	40,1
I mindre grad	49	23,1
Slet ikke	16	7,6
Total	212	100

Kilde: VIVEs spørgeskema til skoleledere.

Bilag 3 Tabeller og figurer til karakteranalyse

I dette bilag fremgår figurer til karakteranalysen i kapitel 2. Analysen ser på fagene:

- Dansk, skriftlig fremstilling
- Matematik med hjælpemidler (9. kl.)/skriftlig matematik (10. kl.)
- Engelsk, skriftlig
- Tysk, skriftlig
- Fransk, skriftlig.

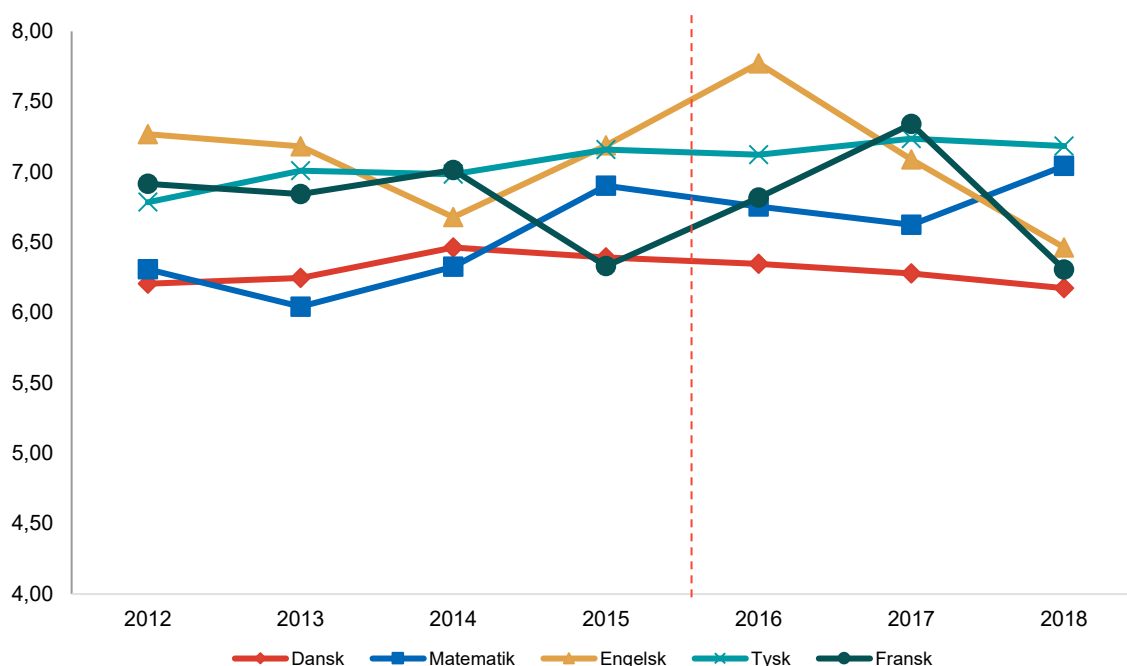
Bilagstabel 3.1 Test af forskelle i udviklingen i skriftlige prøvekarakterer for perioden 2012-2018 med udviklingen for perioden 2016-2018.

	Udvikling 2012-2018	Udvikling 2016-2018	Er udviklingen for 2016-2018 forskellig fra udviklingen 2012-2018
Dansk	-0,01	0,29	Ja
Matematik	0,30	0,17	Ja
Fransk	0,19	-0,12	Ja
Engelsk	0,56	0,66	Nej
Tysk	0,09	0,03	Ja

Anm.: Hausman-test af to skole-klassetrin-fixed-effects modeller, dvs. modeller, hvor der er taget højde for skolen og klassetrinnet. De to modeller adskiller sig ved en variabel, der angiver *tid efter ordningen*. Hausman-testen sammenligner udviklingen i prøvekarakterer henholdsvis med eller uden statistisk kontrol for betydningen af ordningen.

Kilde: VIVE

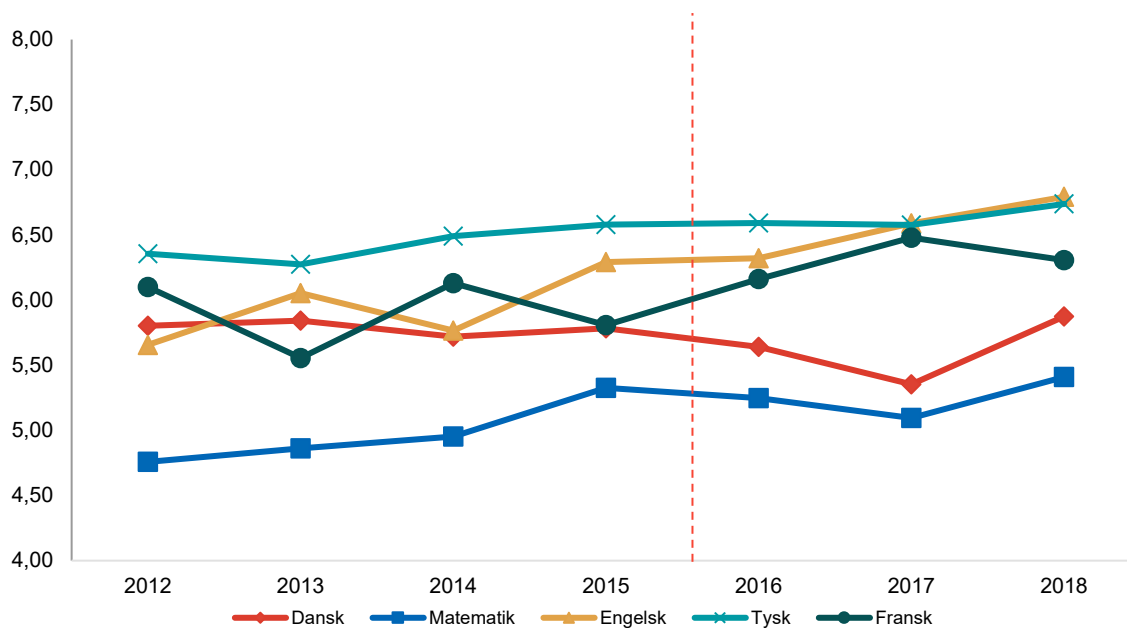
Bilagsfigur 3.1 Den faktiske udvikling i gennemsnitlige prøvekarakterer i skriftlige prøver i 9. klasse. Alle skoletyper. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

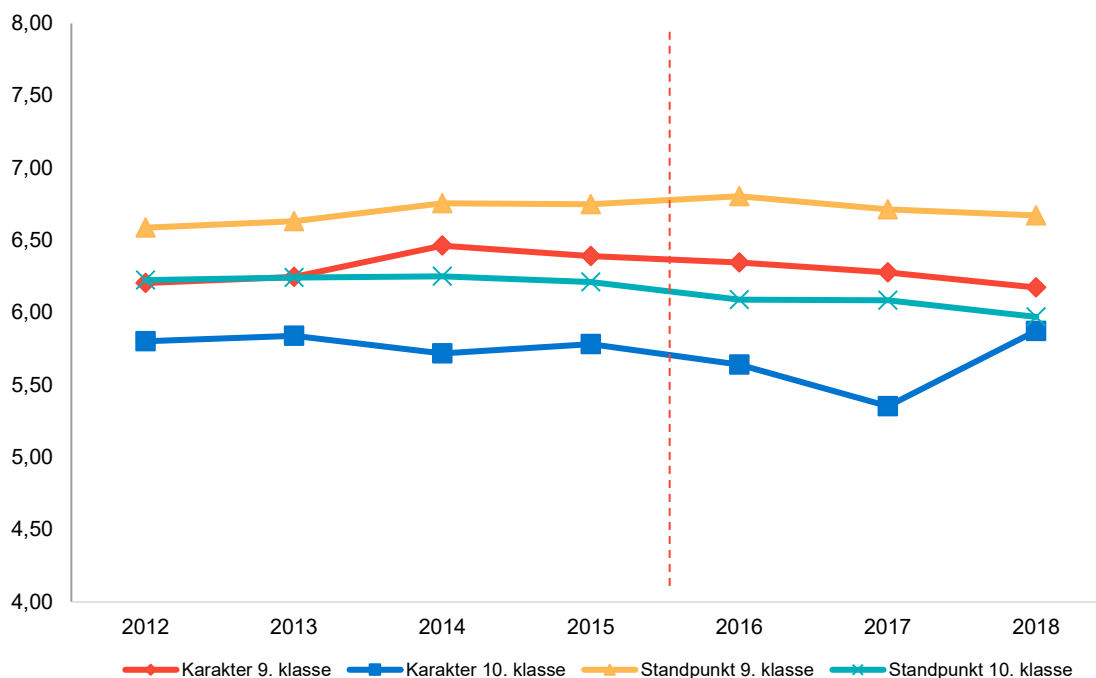
Bilagsfigur 3.2 Den faktiske udvikling i gennemsnitlige prøvekarakterer i skriftlige prøver i 10. klasse. Alle skoletyper. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

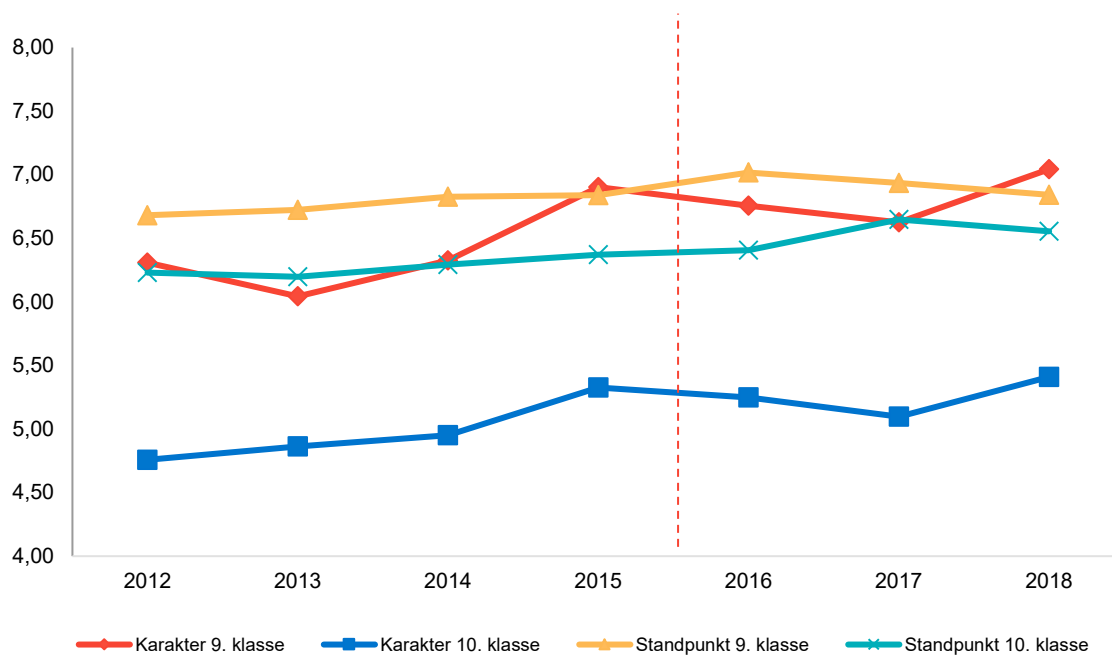
Bilagsfigur 3.3 Faktisk karakterudvikling i dansk. Alle skoler. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

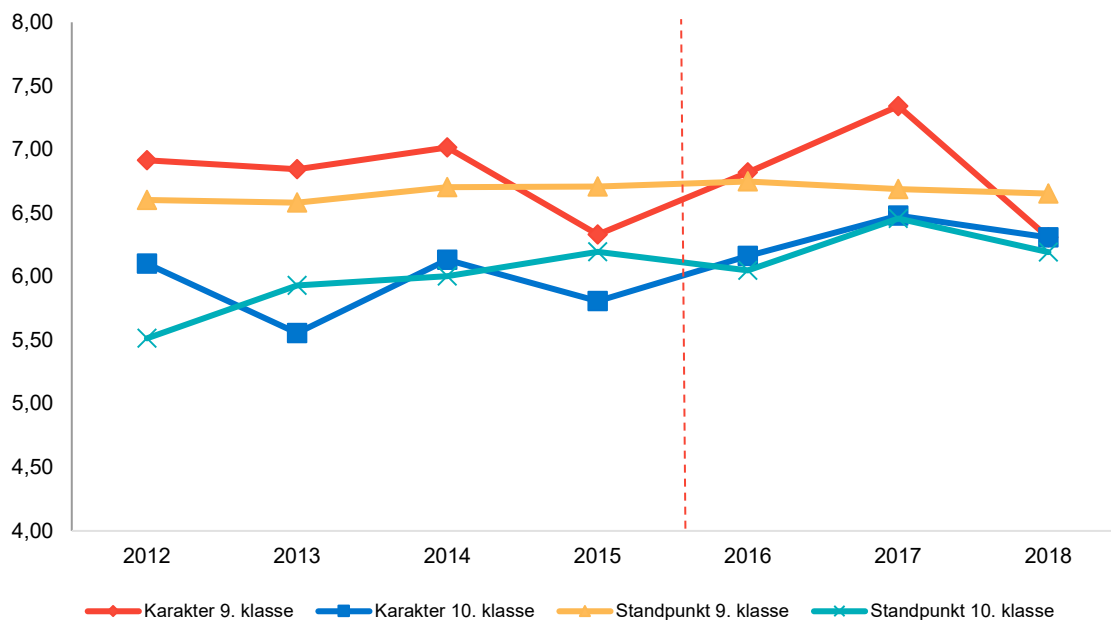
Bilagsfigur 3.4 Faktisk karakterudvikling i matematik. Alle skoler. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

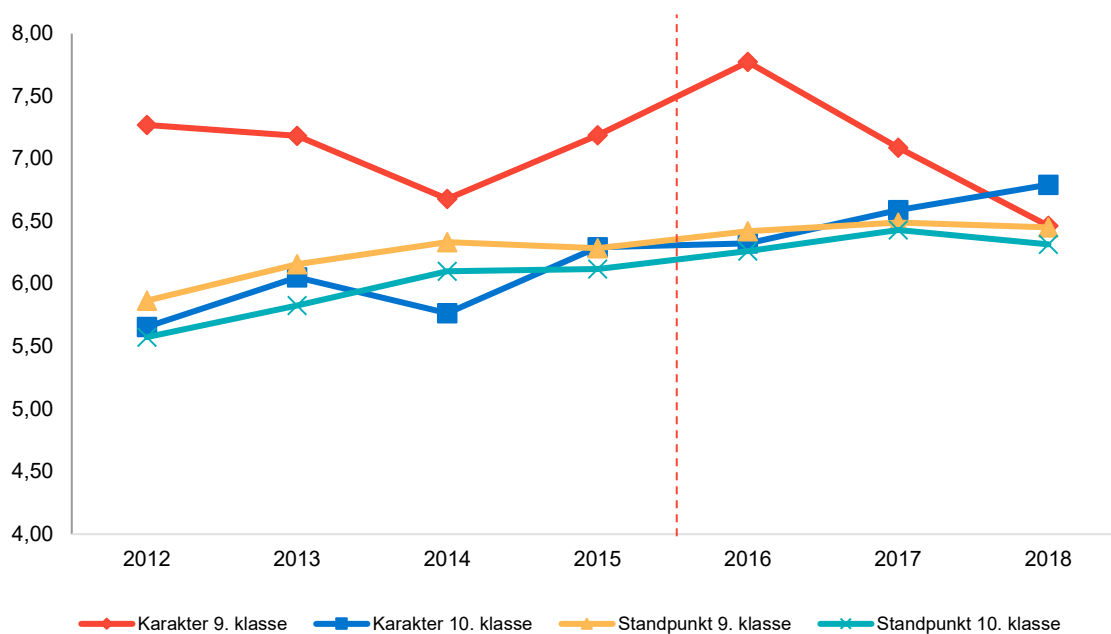
Bilagsfigur 3.5 Faktisk karakterudvikling i fransk. Alle skoler. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

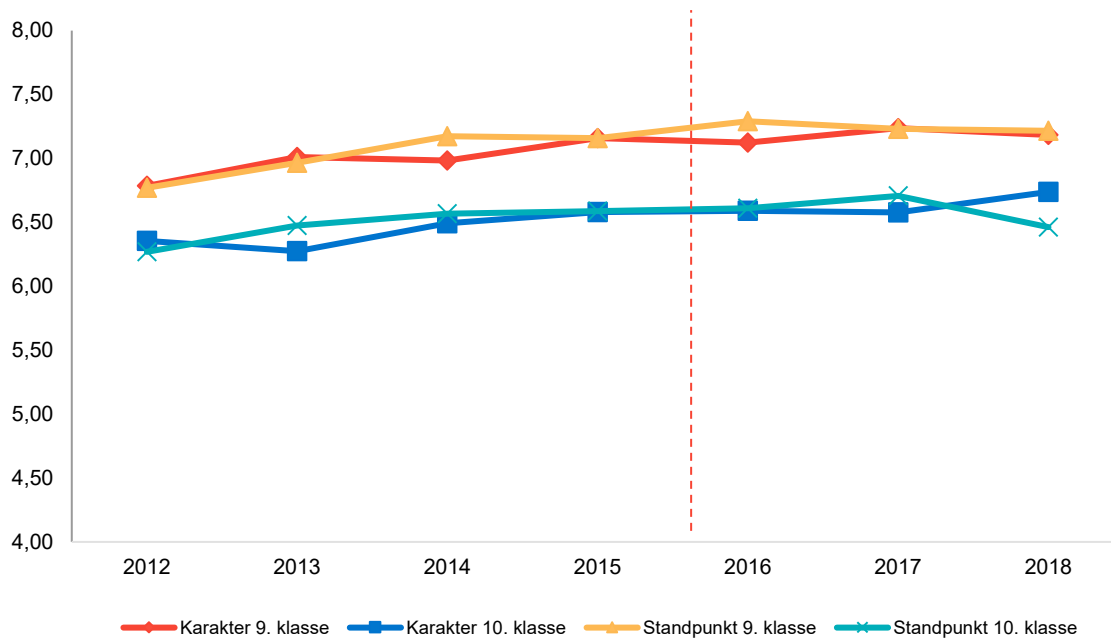
Bilagsfigur 3.6 Faktisk karakterudvikling i engelsk. Alle skoler. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

Bilagsfigur 3.7 Faktisk karakterudvikling i tysk. Alle skoler. 2012-2018.



Anm.: Den lodrette stiplede linje markerer indførelsen af énbedømmerordningen.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

Bilagstabel 3.2 Ændringer i karakterudvikling i karakterpoint. Prøve- og standpunktskarakterer. Alle skoler. 2012-2018. Resultater fra lineær regression med skole-klasse-trin-fixed-effects.

	Dansk	Matematik	Engelsk	Fransk	Tysk
Udvikling i standpunktskarakter over tid	0,02 ** (0,01)	0,01 * (0,01)	0,08 *** (0,01)	0,03 (0,02)	0,11 *** (0,01)
Niveauforskel: prøvekarakterer ift. standpunktskarakterer	-0,32 *** (0,02)	-0,40 *** (0,03)	-0,00 (0,03)	-0,59 *** (0,12)	0,28 *** (0,04)
Ordningens betydning for udviklingen i prøvekarakterniveauet.	0,01 * (0,01)	0,16 *** (0,01)	-0,00 (0,01)	-0,15 *** (0,05)	-0,12 *** (0,02)
Antal observationer	26.164	24.786	19.965	7.751	17.262

Anm.: * = $p < 0,10$, ** = $p < 0,05$, *** = $p < 0,01$. Standardfejl i parentes. Resultaterne er baseret på lineær regression med skole-klasse-trin-fixed effects, med kontrol for trends i udviklingen før ordningen. En falsifikationstest, hvor skæringspunktet er rykket fra 2016 til 2014 (perioden 2016 og fremefter indgår ikke) indikerer, at common trends assumption er overholdt for dansk, engelsk og tysk, men ikke for matematik og fransk.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE.

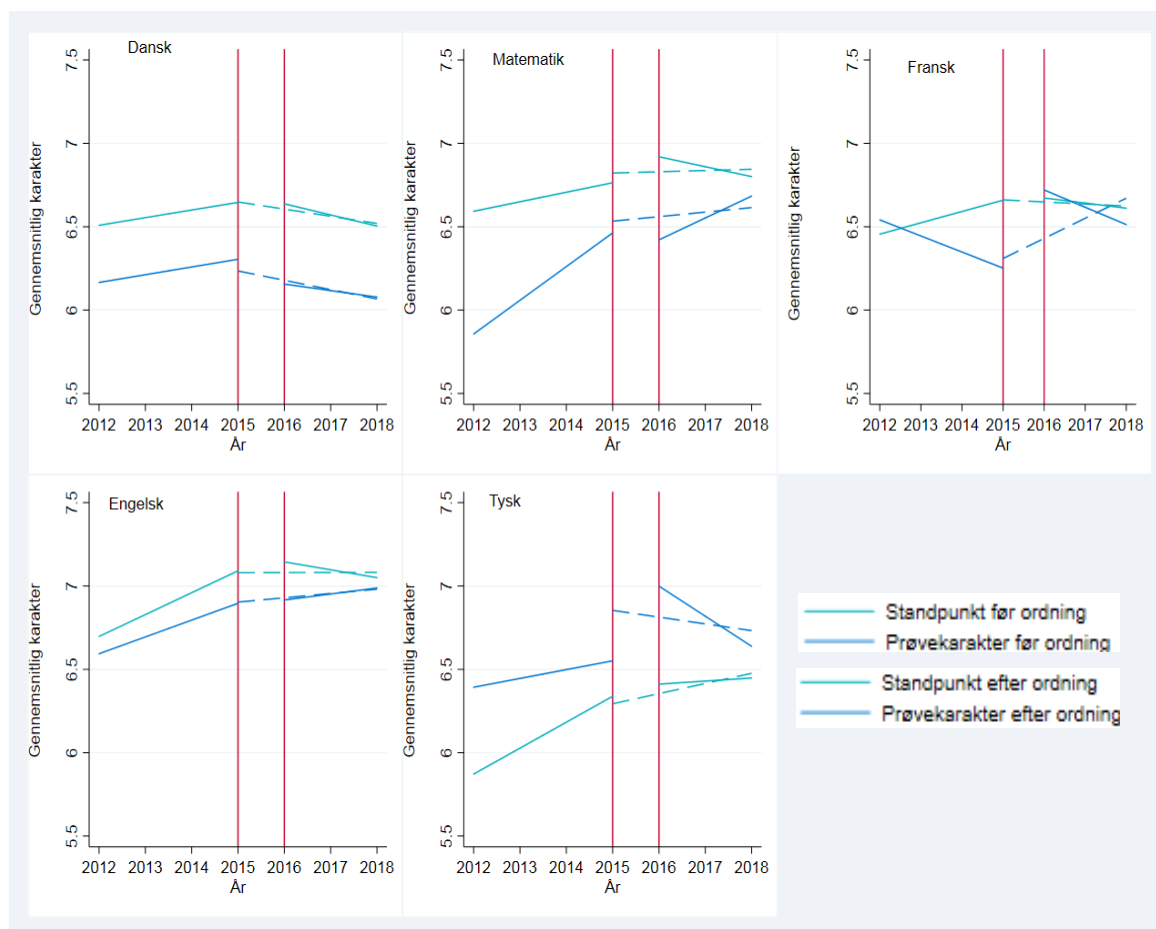
Bilagstabel 3.3 Ændringer i karakterudvikling i karakterpoint. Mundtlige og skriftlige prøvekarakterer. Alle skoler. 2012-2018. Resultater fra lineær regression med skole-klasse-trin-fixed-effects.

	Dansk	Matematik	Engelsk
Udvikling i mundtlige prøvekarakterer over tid	0,02 ** (0,01)	0,18 *** (0,01)	0,05 *** (0,02)
Niveauforskel: skriftlige prøvekarakterer ift. mundtlige prøvekarakterer	-1,45 *** (0,03)	-1,47 *** (0,04)	-0,79 *** (0,07)
Ordningens betydning for udviklingen i det skriftlige prøvekarakterniveau	-0,06 *** (0,01)	0,06 *** (0,01)	-0,04 * (0,02)
Antal observationer	19.907	19.483	9.298

Anm.: * = $p < 0,10$, ** = $p < 0,05$, *** = $p < 0,01$. Standardfejl i parentes. Resultaterne er baseret på lineær regression med skole-klasse-trin-fixed-effects.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE.

Bilagsfigur 3.8 Grafisk repræsentation af Difference-in-Differences analyser for alle fag. Dansk. Matematik. Fransk. Engelsk. Tysk. Lineære regressioner med skoleklassetrin-fixed-effects. 2012-2018.



Anm.: De stiplede linjer repræsenterer forventet udvikling fra 2015-2018. De to lodrette, røde linjer ved hhv. 2015 og 2016 angiver henholdsvis sidste år med to bedømmere (2015) og første år med énbedømmerordningen (2016).

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE.

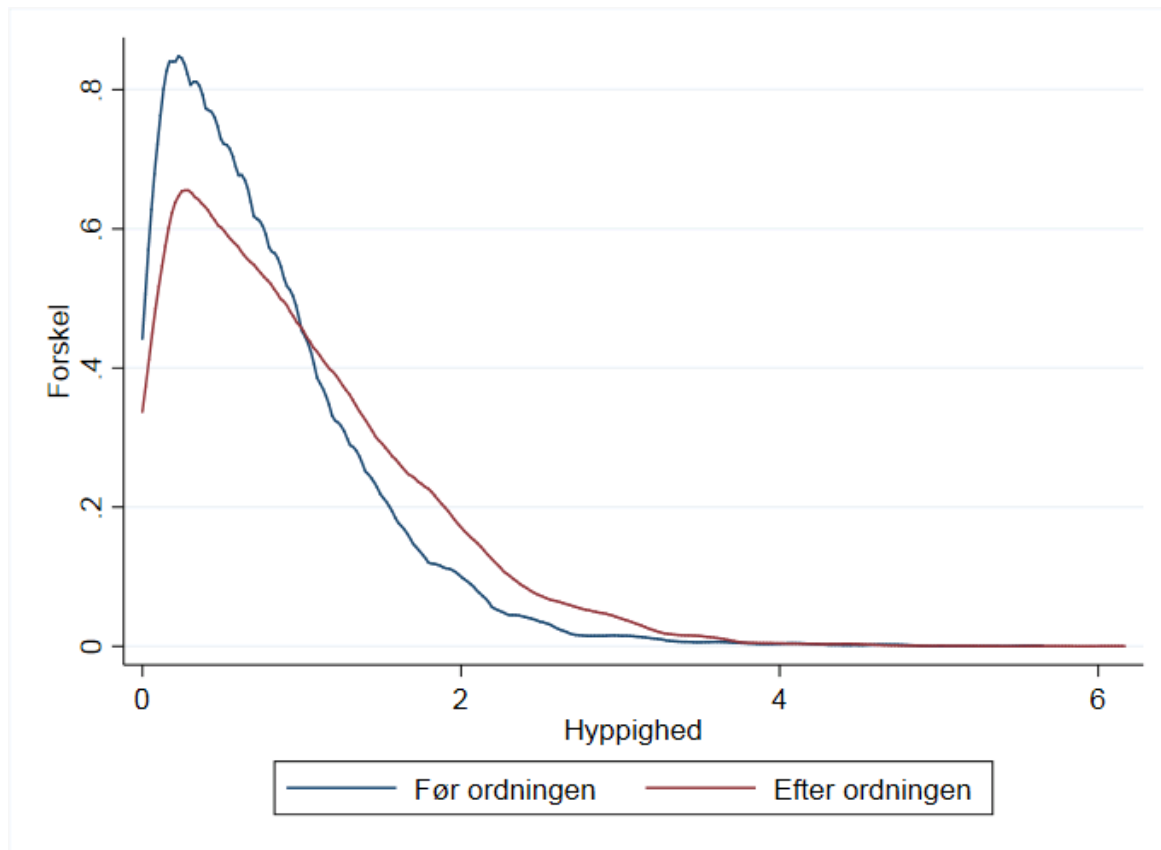
Bilagstabel 3.4 Ændringer i den absolutte forskel mellem prøvekarakterer og standpunktskarakterer Skole-klassetrin-fixed-effects. 2012-2018.

	Dansk	Matematik
Ændring i forskellen mellem prøvekarakterer og standpunkt. 2012-2018	-0,01 (0,01)	-0,03 *** (0,01)
Ændring i forskellen efter indførelsen af ordningen. 2016-2018	0,23 *** (0,02)	0,10 *** (0,02)
Antal observationer	12.888	12.093

Anm.: * = $p < 0,10$, ** = $p < 0,05$, *** = $p < 0,01$. Standardfejl i parentes. Resultaterne er baseret på lineær regression med skoleklassetrin-fixed-effects, med kontrol for trends i udviklingen før ordningen. Målet for forskellen er den absolutte forskel mellem prøvekarakterer og standpunktskarakterer. Et negativt fortegn indikerer, at forskellen bliver mindre, mens et positivt fortegn indikerer, at forskellen bliver større.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger foretaget af VIVE.

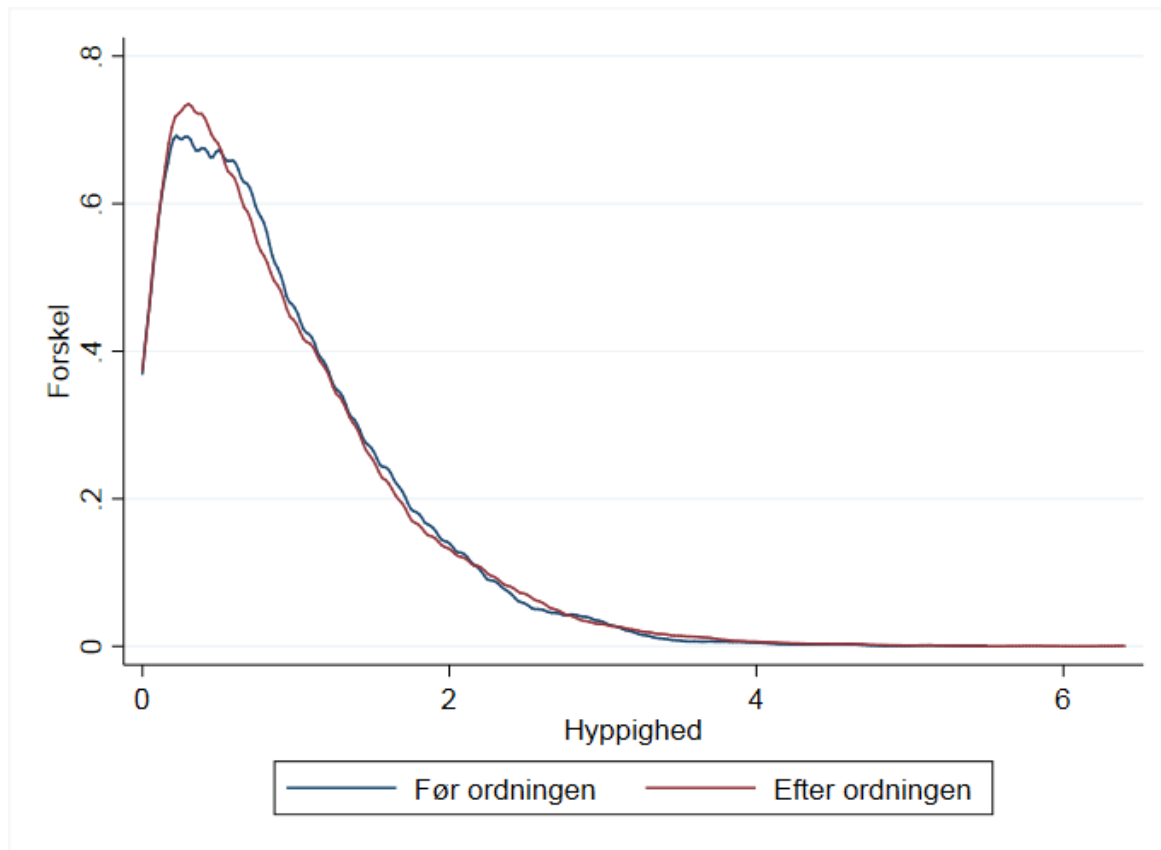
Bilagsfigur 3.9 Fordelinger af forskellen mellem prøvekarakterer og standpunktskarakterer i dansk. Før og efter 2016.



Anm.: Perioden "Før ordningen" dækker over 2012-2015, mens perioden "Efter ordningen" dækker over 2016-2018.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

Bilagsfigur 3.10 Fordelinger af forskellen mellem prøvekarakterer og standpunktskarakterer i matematik. Før og efter 2016.



Anm.: Perioden "Før ordningen" dækker over 2012-2015, mens perioden "Efter ordningen" dækker over 2016-2018.

Kilde: Tal fra Undervisningsministeriet (Styrelsen for IT & Læring). Beregninger fortaget af VIVE.

Litteratur

Angrist, J. D. & J.-S. Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press: Princeton NJ.

Danmarks Evalueringsinstitut (2018) *Bedømmernes perspektiver på 7-trinsskalaen. Bilagsrapport til Evalueringen af 7-trinsskalaen*. Danmarks Evalueringsinstitut: København.

Danmarks Evalueringsinstitut (2019) *Evaluering af 7-trins-skalaen. Samlet rapport*. Danmarks Evalueringsinstitut: København.

Dolin, J., K. Nielsen & B.S. Rangvid (2018) *Rapport fra Følgegruppen for én bedømmer ved folkeskolens prøver – Marts 2018*. [Undervisningsministeriet: København].

**VIDEN I
VELFÆRD**

DET NATIONALE FORSKNINGS-
OG ANALYSECENTER FOR VELFÆRD