

KVALITET OG TILSYN 1

– TILSYNSRAPPORTER SOM VÆRKTØJ I STYRINGEN PÅ DAGTILBUDSOMRÅDET



Indhold

Baggrund	4
Hovedkonklusioner og implikationer	7
Implikationer	8
Overblik over tilsynsrapporter i denne analyse	10
Antallet af rapporter	10
Rapporternes længde	12
Institutionstype	13
KIDS-rapporter	15
Kommunernes KIDS-vurderinger	15
Sammenhæng i KIDS-scorer over tid	16
Sammenholdelse med nationale kvalitetsundersøgelser	17
Sprogværdi som udtryk for kvalitet	22
BOKS 1: Eksempler på sprogværdier	22
Sprogværdi i forhold til kvalitetsvurderinger	22
Sprogværdi som udtryk for andet end kvalitet	23
Narrative rapporter	25
Sprogværdier over tid	25
Sammenhæng i sprogværdier over tid	26
Kommunale forskelle i sprogværdispredning	28
Litteraturliste	32
Bilag	34
Bilag 1. Data- og metodebeskrivelse	34
Bilag 2. Regressionstabel: Persistens i sprogværdier for alle kommuner	40
Bilag 3. Spredning i sprogværdier i alle kommuner vi kan følge over tid	42
Bilag 4. Antal observationer i fordelinger i bilag 3	44
Bilag 5: Fordeling af sprogværdier i 1. og 2. datasæt	45

01

Baggrund

Baggrund

Børns trivsel, læring, udvikling og dannelse i de danske daginstitutioner hænger nøje sammen med kvaliteten. Der er i de senere år blevet foretaget to nationale kvalitetsundersøgelser (EVA, 2020; EVA & VIVE, 2023), og EVA og VIVE er i skrivende stund undervejs med en, der fokuserer på kvaliteten i børnehaver. Fælles for disse undersøgelser er, at de peger på udfordringer med at sikre en ensartet og høj kvalitet i landets daginstitutioner. Andre undersøgelser, der fokuserer på kvaliteten i dagtilbud, viser desuden, at forskellene i kvalitet kan findes på institutions- eller endda stueniveau og altså ikke er koncentreret i bestemte kommuner eller institutionstyper (DEA, 2020, 2021; EVA & VIVE, 2023).

Ansvar for at føre tilsyn med kvaliteten i daginstitutionerne ligger hos de enkelte kommuner, jf. dagtilbudsloven (Dagtilbudsloven, 2019) og de lovmæssige rammer for tilsyn (BUVM, 2020). De kommunale tilsyn skal foretages både anmeldt og uanmeldt af fx pædagogiske konsulenter, som efterfølgende udarbejder tilsynsrapporter for at dokumentere såvel velfungerende praksis som opmærksomhedspunkter, hvor der kan forbedres. Tilsyn med daginstitutioner har været genstand for undersøgelser ved Danmarks Evalueringsinstitut (EVA, 2015), FOA og Bureau 2000 (FOA & Bureau 2000, 2020), samt Tænketanken DEA (DEA, 2023a). Disse undersøgelser finder bl.a., at mens de kommunale tilsyn generelt lever op til de lovmæssige krav, er der stor variation i, hvordan tilsynspraksis ser ud fra kommune til kommune.

Tilsynet har en vigtig intern funktion, der omhandler udviklingsarbejde i de enkelte institutioner (se fx FOA & Bureau 2000, 2020; Københavns Kommune, 2023). Her følges tilsynet op med løbende samtaler og tiltag som fx justeringer af pædagogik praksis eller kompetenceopbygning, hvilket foregår i et samarbejde mellem kommune og institution. Men tilsynet har også en ekstern funktion, hvor kvaliteten i daginstitutionerne via tilsynsrapporterne formidles til fx beslutningstagere og forældre, der ikke altid har adgang til detaljer om det opfølgende arbejde, der foregår internt. Der er en løbende overvejelse med hensyn til, hvordan disse to funktioner i tilsynet balanceres, også fordi de ikke altid stemmer helt overens. Informationer, der kan være vigtige for pædagogisk udviklingsarbejde, er ikke nødvendigvis brugbare for fx forældre. På samme måde er informationer, der kan være vigtige for forældre, ikke nødvendigvis de samme, som kommuner og institutioner lægger vægt på.

Samtidig skal det også påpeges, at tilsynsrapporter for alle aktører blot er én informationskilde blandt andre. Kommunale forvaltninger og til dels beslutningstagere har også adgang til andre data, der kan sige noget om kvalitet, som fx medarbejdertilfredshedsundersøgelser, personaleomsætning, sygefravær og forældretilfredshedsundersøgelser. Forældre kan have adgang til uformelle netværk, egne oplevelser og måske egne observationer i forbindelse med institutionsbesøg. Disse forhold ændrer dog ikke på, at de kommunale tilsyn står som en meget central del af det løbende vurderings- og udviklingsarbejde, når det kommer til kvalitet i daginstitutioner, og derfor er det vigtigt at undersøge dem i deres egen ret.

Som en del af aftalen om minimumsnormeringer fra 2020 kom der nye lovkrav til de kommunale tilsyn. Disse lovkrav omfatter bl.a. et udgangspunkt i det fælles pædagogiske grundlag, sikring af uvildighed, inddragelse af kvantitative og kvalitative data, brug af observationer, opfølgende tiltag, orientering af forældrebestyrelser, vurdering af kvalitet samt offentliggørelse af rapporter. Disse krav trådte i kraft den 1. januar 2022 og skulle være fuldt implementeret senest den 1. maj 2022 (BUVM, 2020). Indtil for nylig har Styrelsen for Udvikling og Kvalitet (STUK) desuden haft et rejsehold, der skulle bistå kommunerne med at implementere de nye lovkrav.

Denne undersøgelse er en opfølgning på DEAs forrige analyse af tilsynsrapporter, der fokuserede på perioden *før*, lovændringen trådte i kraft. I nærværende undersøgelse tilføjer vi rapporter fra tiden *efter* lovændringen for at kunne følge ændringer over tid. Undersøgelsen har flere formål. Dels vil vi gerne undersøge, om de nye tilsynsrammer har ændret på, hvordan kvaliteten af dagtilbud vurderes. Dels vil vi gerne undersøge, hvordan kvalitetsvurderinger adskiller sig fra hinanden, fx i forhold til kommuner, institutionstyper og vurderingsmetoder. Endelig vil vi, mere teknisk, undersøge mulighederne for at anvende tilsynsrapporter som datagrundlag for analyser på et område, hvor det kan være udfordrende at udvikle, operationalisere og følge specifikke parametre for kvalitet. Det skal pointeres, at vi her ser på *vurderingen af kvalitet*, som det kommer til udtryk i tilsynsrapporterne, hvilket adskiller sig fra at se på *kvalitet i sig selv*. Vi kan således kun udtale os om, hvordan tilsynet vurderer kvaliteten, og evt. sammenholde disse vurderinger med andre kvalitetsvurderinger på området. Vi kan ikke ud fra de anvendte data sige noget om det reelle kvalitetsniveau i de danske daginstitutioner eller om forskelle mellem fx institutioner eller kommuner.

Undersøgelsen tager udgangspunkt i følgende undersøgelsesspørgsmål:

1. Hvad kendetegner kommunernes tilsynsrapporter før og efter lovændringen?
2. Hvad vurderes kvaliteten i dagtilbuddene til før og efter lovændringen?
3. Hvilken indvirkning har lovændringen haft på kommunernes tilsynspraksis?

Denne rapport indledes med hovedkonklusioner og implikationer, hvorefter vi analyserer KIDS-rapporter før og efter lovændringen. Det følgende kapitel handler om vores tilgang til de rapporter, der ikke bruger KIDS (det, vi her kalder "narrative rapporter" for at indikere brugen af andre former for strukturer). Disse rapporter analyseres i det efterfølgende kapitel. Rapporten afsluttes med en litteraturliste og bilag, der fremlægger både den anvendte metode og alle de fund, der ligger til grund for undersøgelsen.

02

Hovedkonklusioner og implikationer

Hovedkonklusioner og implikationer

Helt overordnet har vi fundet en række ændringer i forbindelse med lovændringerne og de styrkede krav til de kommunale tilsyn med daginstitutioner. På den ene side er rapporterne blevet længere, og der bliver offentliggjort flere af dem. På den anden side er de overordnede vurderinger stadig meget positive. Disse overordnede tendenser kan brydes ned i en række hovedkonklusioner:

- Der er stor forskel på, hvordan tilsynsrapporterne er udformet. Rapporterne falder i to hovedgrupper. Dels er der KIDS-rapporter, hvor vurderingen af kvalitet formidles via KIDS-scorer, evt. i kombination med en uddybende tekst. Dels er der dem, vi her kalder "narrative rapporter", som dækker over forskellige typer af strukturer, der har det til fælles, at vurderingen af kvalitet primært formidles sprogligt. Især inden for den sidste kategori er der stor variation med hensyn til formater og udformning.
- Antallet af offentliggjorte rapporter stiger. Fra 1. indsamlingsperiode (2018 til 30.4.2022) til 2. periode (1.5.2022 til 2024) er andelen af fundne rapporter steget fra 58 pct. til 76 pct. Dette hænger formentlig sammen med, at et af de nye krav var netop offentliggørelse af rapporter. Samtidig er der forskel på, hvor mange rapporter vi har kunnet finde i hver kommune.
- Kommunernes egne KIDS-scorer er generelt meget positive. Af de 12 kommuner, der i 1. og/eller 2. indsamlingsperiode bruger KIDS-redskabet i deres tilsynsrapporter, vurderer alle på nær én kommune kvaliteten til i gennemsnit at være i spændet fra "god" til "fremragende". Selvom vi ikke direkte kan sammenligne med tidligere nationale kvalitetsundersøgelser, tyder vores resultater på, at kommunernes egne vurderinger tegner et markant mere positivt billede af kvaliteten.
- De narrative rapporter er blevet mere positive efter lovændringen. Det resultat er vi nået frem til ved at benytte en såkaldt "sprogværdi", hvor vi kan kvantificere værdiladningen i den måde, rapporterne er formuleret på. Samtidig er der i den gennemsnitlige spredning reduceret, så afstanden mellem de bedste vurderinger og de dårligste er blevet mindre. Ændringerne i spredning vidner om, at mange kommuner har ændret tilsynsmodel fra 1. til 2. indsamlingsperiode.
- Der er en gruppe af institutioner, der konsekvent vurderes negativt i både 1. og 2. indsamlingsperiode. Selvom der generelt er stor mobilitet med hensyn til kvalitetsvurderinger, så en institution med god eller mellemgod vurdering i 1. indsamlingsperiode ikke nødvendigvis opnår samme resultat i 2. indsamlingsperiode, er der en undtagelse. Af de institutioner, der i 1. indsamlingsperiode ligger i den laveste fjerdedel inden for hver kommune, beholder 35 pct. denne position i 2. periode. Det tyder på, at der er en gruppe af institutioner, hvor det ikke lykkes at forbedre vurderingerne af kvaliteten.

Implikationer

Vores resultater tyder på, at de nye tilsynskrav i forbindelse med lovændringen i 2022 har haft en indvirkning på tilsynspraksis og tilsynsrapporter i kommunerne. Fx bliver der offentliggjort flere rapporter, og rapporterne er blevet mere omfangsrige. Der er dog stadig enkelte steder, hvor man kunne forbedre praksis i forbindelse med offentliggørelse af rapporter og præcisering af vurderingerne. Samtidig kan det være en fordel at arbejde bevidst med, hvordan vurderings- og udviklingsperspektiverne i tilsynsrapporterne skal balanceres, så både forvaltninger, beslutningstagere og forældre har gavn af informationerne.

Det kræver dog stadig, at de tilgængelige informationer om kvaliteten er retvisende. Her er den markante forskel mellem, hvordan kommunerne selv vurderer kvaliteten i deres daginstitutioner, og hvordan institutionerne vurderes i de nationale kvalitetsundersøgelser, bemærkelsesværdig. Ud fra de her anvendte data er det vanskeligt at afgøre de præcise årsager til denne tendens, men for at komme det nærmere peger vi på følgende tre opmærksomhedspunkter i forhold til at arbejde videre med problemstillingen.

For det første er der spørgsmålet om redskaber. Selvom fx KIDS er et standardiseret værktøj, tyder vores resultater på, at det kan have betydning, hvem der bruger det. Det kalder på en opmærksomhed i forhold til fx KIDS-scorer og hvordan man sikrer, at de anvendte redskaber giver en reel indsigt i kvaliteten i de vurderede daginstitutioner.

For det andet er der spørgsmålet om uvildighed, som er et af kravene i den nye lovgivning om tilsyn med dagtilbud. Der kan være mange gode grunde til, at det i nogle tilfælde er vanskeligt at sikre et fuldstændig uvildigt tilsyn, heriblandt kommunale ressourcer, interne processer og personlige relationer. Men netop derfor er det essentielt at være opmærksom på, hvordan man skaber vilkårene for at udarbejde uvildige vurderinger.

For det tredje er der spørgsmålet om, hvordan man sikrer et kvalitetsløft af de institutioner, der har behov for det. Hvis en stor del af institutionerne i den nederste fjerdedel forbliver i denne kategori over tid, kalder det på at forbedre grundlaget for målrettede indsatser, der kan forbedre kvaliteten.

Endelig peger undersøgelsen på, at der er markante forskelle mellem kvalitetsvurderinger, både mellem kommuner og over tid. Dette gælder også, selvom der anvendes det samme standardiserede redskab, som fx KIDS. Set i det lys kunne det være en fordel at have en uafhængig instans, som kan facilitere vidensdeling, være opsøgende og samtidig støtte op om implementeringen af god vurderingspraksis på tværs af kommuner. Denne funktion kan enten være statslig eller ligge i kommunalt regi, det centrale er at sikre, at kommunernes vurdering af kvalitet ligger så tæt på den reelle kvalitet som muligt.

03

Overblik over tilsynsrapporter i denne analyse

Overblik over tilsynsrapporter i denne analyse

Antallet af rapporter

Vi har fundet 3.079 tilsynsrapporter fra perioden 1.5.2022 til 2024. Det svarer til rapporter fra 76 pct. af landets daginstitutioner,¹ hvor alle landets 98 kommuner er repræsenteret. Efter bortsortering af rapporter, der ikke lader sig analysere (fx fordi de er håndskrevne, scannede eller tomme), er antallet af brugbare rapporter 2.896. Vi refererer til denne samling af tilsynsrapporter som '2. datasæt' eller 'det nye datasæt'.

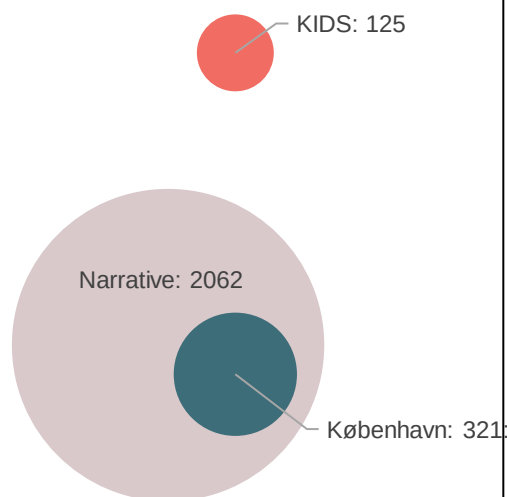
Vi sammenligner løbende de 2.896 tilsynsrapporter med en tidligere indsamlingsperiode af tilsynsrapporter fra perioden 2018 til 30.4.2022. Dette datasæt kalder vi i denne rapport for '1. datasæt' eller 'det gamle datasæt', og det har tidligere været genstand for analyse (DEA, 2023a). Dette 1. datasæt består af 2.289 tilsynsrapporter fra 83 kommuner, hvilket svarer til 58 pct. af landets daginstitutioner. Der er altså sket en stigning i antallet af offentliggjorte rapporter, hvilket også er forventeligt, eftersom der siden 2022 har været et lovkrav om offentliggørelse af tilsynsrapporter som en del af aftalen om minimumsnormeringer (BUVM, 2020).

I analysen arbejder vi med tilsynsrapporterne opdelt i to hovedtyper: rapporter rapporteret med KIDS-scorer og det, vi her kalder for 'narrative rapporter'. Derudover har Københavns Kommune sit eget tilsynskoncept (Københavns Kommune, 2023), som pga. antallet af daginstitutioner i kommunen fylder en del i det samlede datasæt. Antallet af de forskellige typer af kategorier fra både 1. og 2. indsamlingsperiode er illustreret i figur 1.

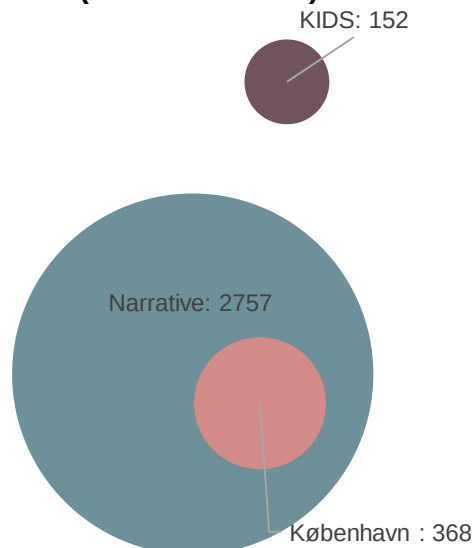
¹ I forhold til STILs dagtilbudsregister.

Figur 1 // Sammensætning af rapporter i de to dataindsamlinger

1. datasæt (2018 til 30.4.2022)



2. datasæt (1.5.2022 til 2024)



Datasættene inkluderer både kommunale, selvejende og private daginstitutioner. En del af de 2.896 rapporter fra 2. datasæt (dvs. perioden 1.5.2022 til 2024) beskriver nyoprettede institutioner eller institutioner, vi ikke tidligere har kunnet finde tilsynsrapporter for. Vi kan identificere 1.417 institutioner, der er med i begge datasæt. Det svarer til halvdelen af institutionerne i 2. datasæt.

Der er en vis forskel på, hvor mange tilsynsrapporter vi har kunnet finde i hver kommune. 17 kommuner har en meget høj grad af tilgængelighed med fundne rapporter for over 95 pct. af kommunens institutioner. For lidt over halvdelen af kommunerne (55 pct.) finder vi rapporter for over 80 pct. af kommunens institutioner. I den anden ende af skalaen er der 7 kommuner, hvor vi kun har kunnet finde rapporter for maks. en tredjedel af kommunens institutioner. I 6 tilfælde har vi fundet under fem offentliggjorte rapporter for hele kommunen².

Der er imidlertid en vis usikkerhed i forhold til, om rapporterne ikke er offentliggjort, eller om vi blot ikke har kunnet finde dem. Vi kan ikke afvise, at der er tilfælde, hvor vi ikke har fundet ellers offentliggjorte tilsynsrapporter. Dog er hele formålet med offentliggørelsen af tilsynsrapporter, at det skal bidrage til åbenhed og gennemsigtighed om kvaliteten (BUVM, 2020), så hvis det ikke er muligt at identificere rapporterne gennem en systematisk søgning, udgør det stadig et problem i forhold til dette formål.

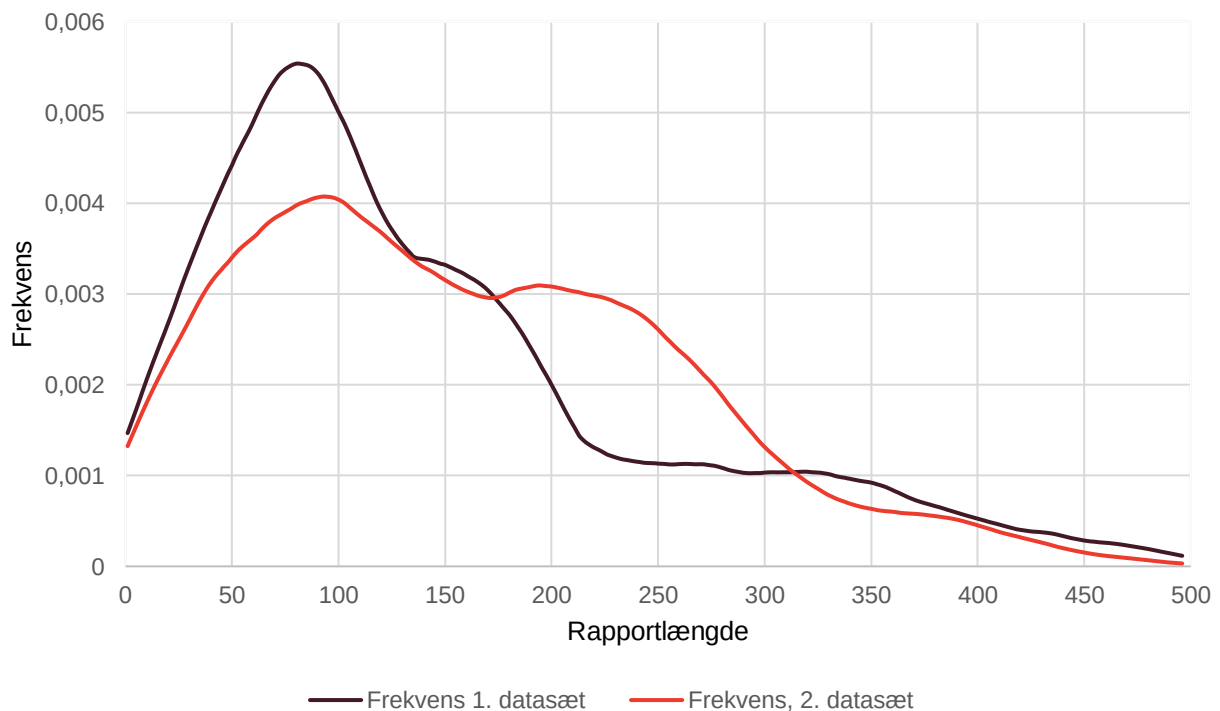
² Den fulde oversigt er tilgængelig i bilag 1.

Det skal også bemærkes, at denne optælling ikke forholder sig til indholdet af rapporterne. Mens nogle rapporter er meget udførlige med både kvantitative og kvalitative data (jf. aftalen om minimumsnormeringer), er andre mere ordknappe, fx ved kun at indeholde én sætning om, at tilsynet er gennemført uden anmærkninger. Det skal dog bemærkes, at lovkravet om offentliggørelse ikke specificerer, hvilken form tilsynsrapporterne skal have, og at en enkelt sætning reelt lever op til lovens bogstav. Dog vil informationsværdien for eksempelvis forældre være begrænset.

Rapporternes længde

Sammenlignet med 1. datasæt er rapporterne i gennemsnit blevet længere. Dette kan afspejle en opfyldelse af intentionen med det styrkede tilsyn, nemlig at sikre grundigere tilsyn. Som det ses i figur 2, havde mange tilsynsrapporter fra 1. datasæt omtrent samme længde på omkring 80 sætninger, mens spredningen i 2. datasæt er blevet betydeligt større. Der er nu en lille top omkring 96 sætninger og en generelt fladere fordeling med betydeligt flere rapporter med over 200 sætninger.

Figur 2 // Længde af rapporter fra 1. til 2. datasæt



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024

Anm.: $n = 2.125$ og 2.890 for 1. og 2. datasæt, hhv. Længde er defineret som antallet af sætninger i rapporten. Rapporter med flere end 500 sætninger eller en sprogverdi (se senere kapitel herom) større end 7 er ikke vist.

Samtidig er der stor forskel på, hvordan kommunerne griber formatet for rapporter an. Dette er helt efter bogen, eftersom lovteksten om styrket tilsyn ikke fastsætter nogen fast struktur for rapporterne (BUVM, 2020). Mens nogle kommuner bruger et redskab som KIDS (Ringsmose & Kragh-Müller, 2020), har andre kommuner udarbejdet deres egne formater, som kun bruges i den pågældende kommune. Og mens de fleste

kommuner udarbejder tilsynsrapporter for hver af deres daginstitutioner, lader andre kommuner hver rapport dækker alle institutioner i fx en klynge.³

Institutionstype

Der er ikke signifikant forskel på andelen af institutioner, vi har fundet rapporter for, i det nye datasæt på tværs af institutionstyper, jf. tabel 1. Tværtimod afviger andelen af fundne rapporter kun et par procentpoint fra den overordnede andel på 76 pct., hvad angår både børnehaver, vuggestuer og integrerede institutioner. Generelt er der en klar overvægt af integrerede institutioner, både i den samlede mængde institutioner og i vores datasæt.

Tabel 1 // Overblik over institutionstyper

	Antal institutioner i alt	Antal rapporter fundet	Pct. fundet	Gennemsnitlig længde
Alle typer	4.058	3.079	75,9	159,0
Integreret	2.921	2.257	77,3	158,1
Børnhave	881	630	71,5	157,6
Vuggestue	256	192	75,0	174,4 ^{**}

Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioden 1.5.2022 til 2024

Anm.: 'Længde' er defineret som antal sætninger i en rapport. Vi har ikke data på institutionstype for første dataindsamling. Den gennemsnitlige længde for vuggestuerne er signifikant længere end for de to andre institutionstyper på et 5-pct.-signifikansniveau. Den gennemsnitlige længde er udregnet på baggrund af de brugbare rapporter, mens kolonnen "Antal rapporter fundet" refererer til den samlede mængde rapporter fundet, inklusive ikke-brugbare rapporter. Der indgår hhv. n = 2.889, n = 2.107, n = 599 og n = 183 rapporter i de fire gennemsnit.

Der er dog forskel på den gennemsnitlige rapportlængde, idet tilsynsrapporterne for vuggestuer er en del længere (174 sætninger), end for de øvrige institutionstyper (lige under 160 sætninger), som det fremgår af tabel 1.

³ Da det ikke er muligt at sammenligne klynge-rapporter med institutionsspecifikke rapporter, er de udeladt i denne undersøgelse.

04

KIDS-rapporter

KIDS-rapporter

I dette kapitel ser vi nærmere på de tilsynsrapporter, der benytter sig af KIDS, Kvalitetsudvikling i daginstitutioner, som er et danskudviklet redskab til kvalitetsvurdering af læringsmiljøer (Ringsmose & Kragh-Müller, 2020). KIDS baserer sig på observationer inden for de tre overordnede områder "fysiske omgivelser", "relationer" samt "leg og aktivitet", samt seks specifikke områder. Inden for hver kategori er der en række spørgsmål, der tildeles scorer af observatørerne. Det resulterer dels i en score for hvert område, dels i en samlet kvalitetsscore. Denne kvantificering af observationer i KIDS muliggør sammenligninger på tværs af daginstitutioner⁴. Samtidig kan de samlede scorer inddeles i hhv. "fremragende kvalitet", "god kvalitet", "tilstrækkelig kvalitet", "utilstrækkelig kvalitet" eller "ringe kvalitet", sådan som også EVA og VIVE gjorde i forbindelse med den nationale kvalitetsundersøgelse af vuggestuer og dagplejer (EVA & VIVE, 2023).

Det skal bemærkes, at en kommune i princippet godt kan benytte sig af KIDS uden at anvende den kvantitative scoringsmulighed eller uden at lade resultaterne indgå i en tilsynsrapport. Det kan være tilfældet, enten hvis kommunen bruger KIDS i andre sammenhænge end i tilsynet (fx til løbende kvalitetsudvikling ud over tilsyn), eller hvis kommunen bruger KIDS i tilsynet uden at offentliggøre scorer i tilsynsrapporterne. Begge dele er helt i tråd med lovgivningen på området, men det bevirker, at vi ikke via tilsynsrapporterne har haft adgang til specifikke scorer.

Kommunernes KIDS-vurderinger

I det følgende fokuserer vi på de kommuner, der både anvender KIDS direkte i deres tilsyn og offentliggør institutionernes KIDS-scorer i tilsynsrapporterne. Det drejer sig om 125 rapporter i 1. datasæt og 152 rapporter 2. datasæt, som det fremgår af tabel 2. Tabellen viser også den gennemsnitlige KIDS-score⁵, og antallet af rapporter med KIDS-scorer for de kommuner, der benytter værktøjet.

Tabel 2 // Kommuner, der anvender KIDS

Kommune	1. datasæt			2. datasæt		
	Gns. KIDS-score	Gns. kvalitets-kategori	Antal KIDS-rapporter	Gns. KIDS-score	Gns. kvalitets-kategori	Antal KIDS-rapporter
Alle rapporter	70,2	God	125	68,5	God	152
Sønderborg	64,8	God	46	63,7	God	38
Herlev	70,1	God	16	73,8	God	22
Slagelse	67,9	God	14	73,1	God	41
Allerød	66,6	God	7	70,7	God	9
Brøndby	72,4	God	21	.	.	.
Frederikssund	87,8	Fremragende	16	.	.	.

⁴ Vi sammenligner med forbehold for, at personalet, der foretager kvalitetsvurderingerne, ikke er det samme på tværs af de kommuner, vi sammenligner. Se det separate metodeafsnit for uddybning af antagelserne bag dette.

⁵ Idet kommunerne ikke alle anvender alle spørgsmål i KIDS, definerer vi en vægtet KIDS-score for hver institution, $S_w = (S - L) / (H - L)$, hvor S er den samlede score, L og H er den hhv. lavest og højest mulige score. S_w kan sammenlignes på tværs af institutioner, hvor L og H ikke er ens. Den vægtede score tager højde for de forskellige antal spørgsmål, der besvares i rapporterne.

Viborg	61,5	Tilstrækkelig	3	.	.	.
Lolland	65,7	God	1	.	.	.
Roskilde	82,9	God	1	.	.	.
Kolding	.	.	.	63,4	God	39
Syddjurs	.	.	.	96,1	Fremragende	2
Ærø	.	.	.	69,6	God	1

Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024

Anm.: Her ser vi bort fra alle rapporter, der ikke benytter det fulde KIDS-værktøj. Det inkluderer rapporter, der skriver, at de benytter KIDS, men som ikke angiver de opnåede KIDS-scorer. Den gennemsnitlige kvalitetskategori korresponderer til den kvalitetskategori, der fremgår af KIDS-værktøjet på skalaen "ringe", "utilstrækkelig", "tilstrækkelig", "god" og "fremragende" kvalitet. Se metodebilaget for en beskrivelse af både kvalitetskategorien og den gennemsnitlige KIDS-score.

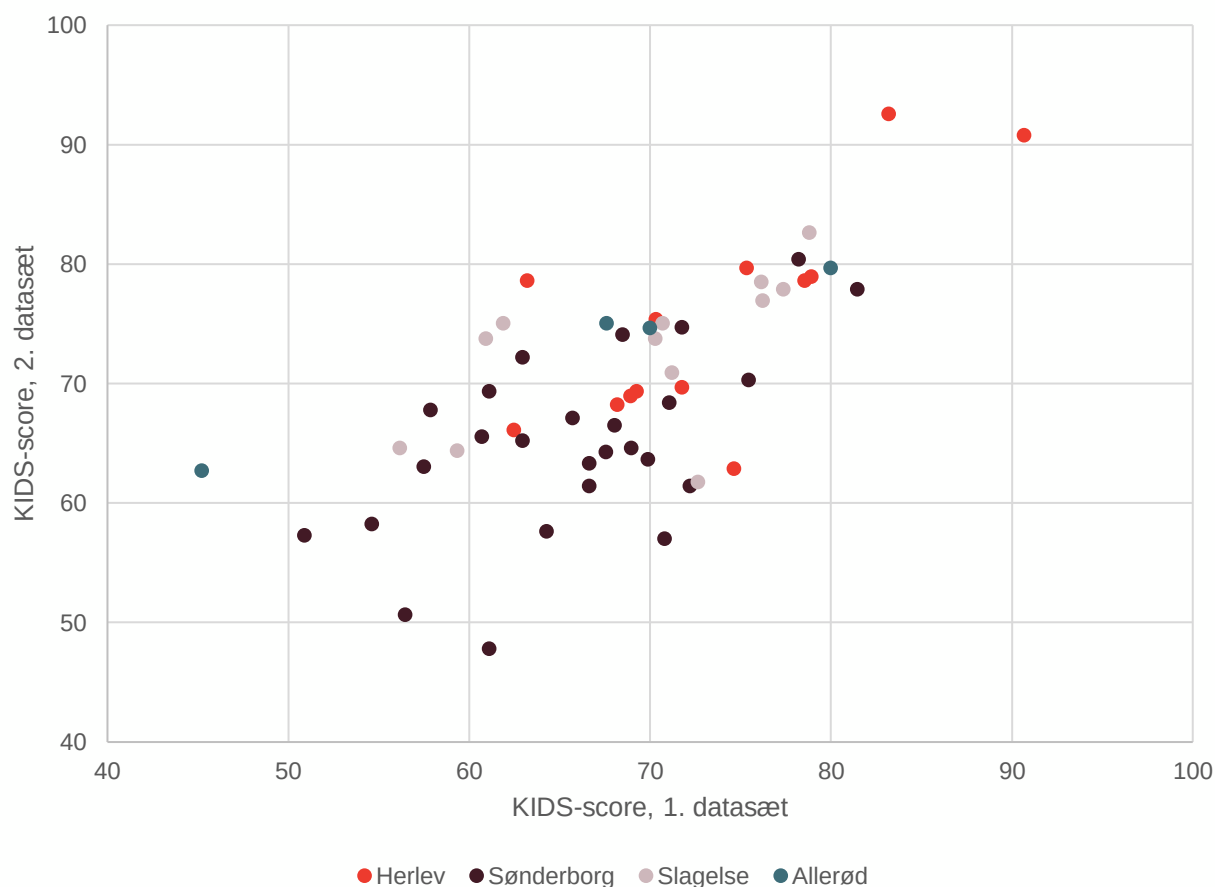
Som det fremgår af første række i tabel 2, er den gennemsnitlige KIDS-score i både 1. og 2. datasæt i kategorien "god" (hhv. 70,2 og 68,5 point). Det betyder, at der trods variation i forhold til de enkelte institutioners KIDS-scorer og variation i forhold til, hvilke elementer fra KIDS tilsynene anvender, er en generel tendens til, at institutionernes kvalitet vurderes meget positivt. Kigger man på de enkelte kommuners KIDS-scorer, ses det også, at der kun er ét tilfælde, hvor den gennemsnitlige kvalitet vurderes som "tilstrækkelig" (Viborg i 1. datasæt, hvor tre institutioner vurderes). Ellers ligger de gennemsnitlige vurderinger på tværs af datasæt i spændet fra "god" til "fremragende".

Sammenhæng i KIDS-scorer over tid

Som det ses øverst i tabel 2 er der fire kommuner, der har brugt KIDS i begge indsamlingsperioder (Sønderborg, Herlev, Slagelse og Allerød). Derudover er der fem kommuner, der kun har brugt KIDS i 1. datasæt (Brøndby, Frederikssund, Viborg, Lolland og Roskilde), mens der er tre kommuner, der kun har brugt KIDS i 2. datasæt (Kolding, Syddjurs og Ærø). Selvom denne varians vanskeliggør en systematisk sammenligning af udviklingen i KIDS-scorer over tid på kommuneniveau, kan vi spore visse tendenser i de 55 institutioner fra de fire kommuner, der har anvendt KIDS i begge datasæt.

Figur 3 viser således, at de institutioner, der i 1. datasæt havde en høj KIDS-score, også scorer højt i 2. datasæt. I alle fire kommuner er der en stærk positiv sammenhæng mellem scorerne i de to perioder, hvilket indikerer en stærk sammenhæng i vurderingen af kvaliteten over tid. Alt i alt kan omkring 37 pct. af variansen i de nye KIDS-scorer forklares af de gamle KIDS-scorer. Det betyder også, at tilsynene vurderer, at de daginstitutioner der fik en relativt lav score i 1. datasæt, ikke har forbedret sig betydeligt i næste periode.

Figur 3 // KIDS-scorer i 1. og 2. datasæt



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024

Anm.: Vi definerer en vægtet KIDS-score for hver institution, $S_w = (S - L)/(H - L)$, hvor S er den samlede score, L og H er den hhv. lavest og højest mulige score. S_w kan sammenlignes på tværs af institutioner, hvor L og H ikke er ens.

Den stærke sammenhæng i vurderingen af kvalitet over tid, der ses i figur 3, reflekterer resultater fra tidligere undersøgelser. Fx viser undersøgelser af medarbejdertilfredshed blandt pædagogisk personale (DEA, 2023b) og daginstitutionernes betydning for børnenes senere faglige resultater (DEA, 2024) samme tendens til sammenhæng over tid. Denne sammenhæng over tid kan tolkes på den måde, at det i mange tilfælde ikke lykkes at adressere uhensigtsmæssige mønstre, hvad enten det drejer sig om KIDS-scorer, medarbejdertilfredshed eller børnenes faglige udbytte, også selvom man er bekendt med udfordringerne.

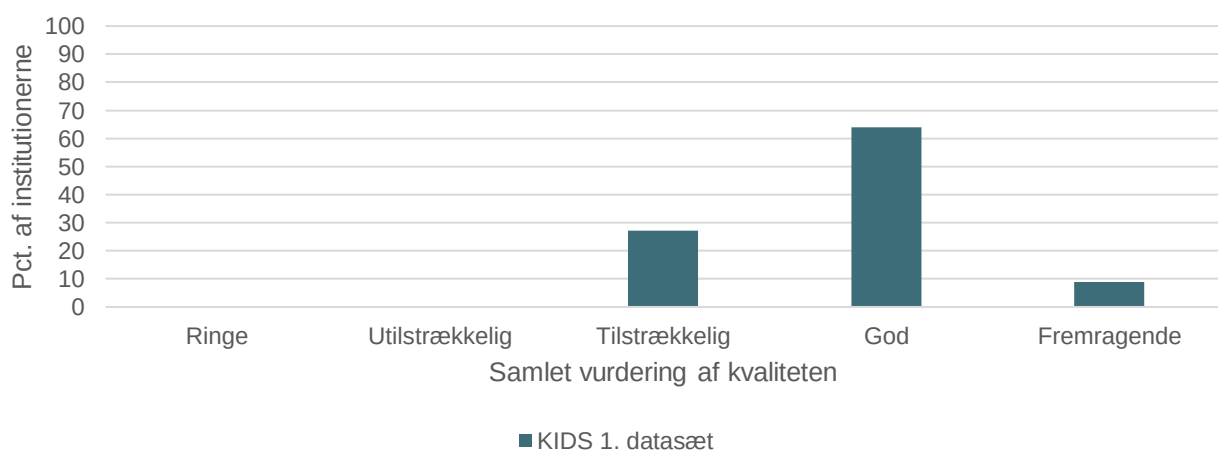
Sammenholdelse med nationale kvalitetsundersøgelser

Der er i 2020 og 2023 foretaget to nationale kvalitetsundersøgelser i dagtilbud af hhv. EVA og EVA/VIVE (EVA, 2020; EVA & VIVE, 2023). Den første undersøgelse benyttede sig af redskabet ECERS-3 til at vurdere kvaliteten i børnehaver (Harms et al., 2014), hvilket affødte en livlig debat om kvalitetsmarkører og måling af kvalitet i dagtilbud (EVA, 2022). Den anden undersøgelse fokuserede på vuggestuer og dagplejer og benyttede sig af redskabet KIDS (EVA & VIVE, 2023). I skrivende stund arbejder EVA og VIVE på anden del af denne undersøgelse, der fokuserer på børnehaver.

Vi kan i nærværende undersøgelse ikke direkte sammenligne resultaterne fra kommunernes KIDS-vurderinger med resultaterne fra de nationale kvalitetsundersøgelser. For det første er der forskelle mellem ECERS-3 og KIDS. For det andet er der forskelle i måden, hvorpå EVA/VIVE operationaliserer KIDS, sammenholdt med kommunerne. For det tredje fokuserer EVA og EVA/VIVES undersøgelser på hhv. vuggestuer/dagplejer og børnehaver mens vores inkluderer alle typer. Og for det fjerde er der relativt få institutioner, der i vores undersøgelse er blevet vurderet med KIDS i både 1. og 2. datasæt, hvilket giver en vis usikkerhedsmargin. På trods af dette er der dog visse tendenser, der træder frem, når man sammenholder resultaterne uden direkte at sammenligne dem.

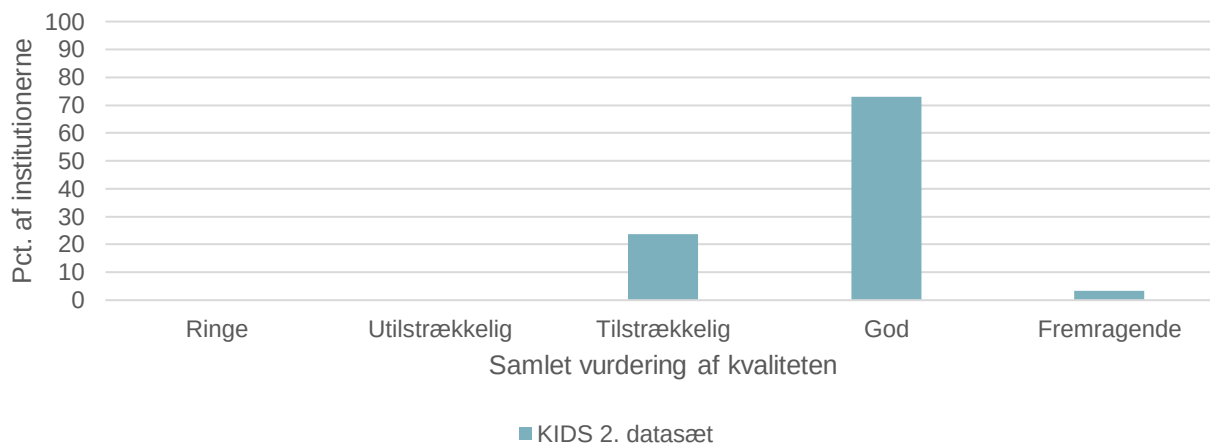
De KIDS-vurderinger, vi i nærværende undersøgelse finder i både 1. og 2. datasæt, er generelt meget positive, som det fremgår af figur 4 og 5. Her ses det, at vurderingerne på tværs af kommuner og institutioner centrerer sig omkring "god" (64 pct. i 1. datasæt og 73 pct. i 2. datasæt), mens der er lidt færre i kategorierne "tilstrækkelig" (27 pct. i 1. datasæt og 24 pct. i 2. datasæt) og "fremragende" (9 pct. i 1. datasæt og 3 pct. i 2. datasæt). I ingen af datasættene er der nogen institutioner, der er vurderet som "utilstrækkelig" eller "ringe".

Figur 4 // Fordelingen af KIDS-scorer i 1. datasæt



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioden 2018 til 30.4.2022
Anm.: I figuren er KIDS-scorer fortolket som kvalitetskategorier (EVA & VIVE, 2023).

Figur 5 // Fordelingen af KIDS-scorer i 2. datasæt



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioden 1/5/2022 til 2024

Anm.: I figuren er KIDS-scorer fortolket som kvalitetskategorier, jf. EVA/VIVES metode (EVA & VIVE, 2023).

Vender vi blikket mod EVA og VIVES repræsentative undersøgelser fra 2020 og 2023 er givet de et andet indtryk. Som det fremgår af figur 6 og 7 centrerer disse vurderinger sig omkring "tilstrækkelig" (75 pct. i EVAs i 2020 og 49 pct. i EVA/VIVES i 2023). Dertil kommer kategorierne "utilstrækkelig" (6 pct. i EVAs i 2020 og 38 pct. i EVA/VIVES i 2023) og "god" (19 pct. i EVAs i 2020 og 13 pct. i EVA/VIVES i 2023). I ingen af de to undersøgelser er der nogen institutioner i kategorien "fremragende".

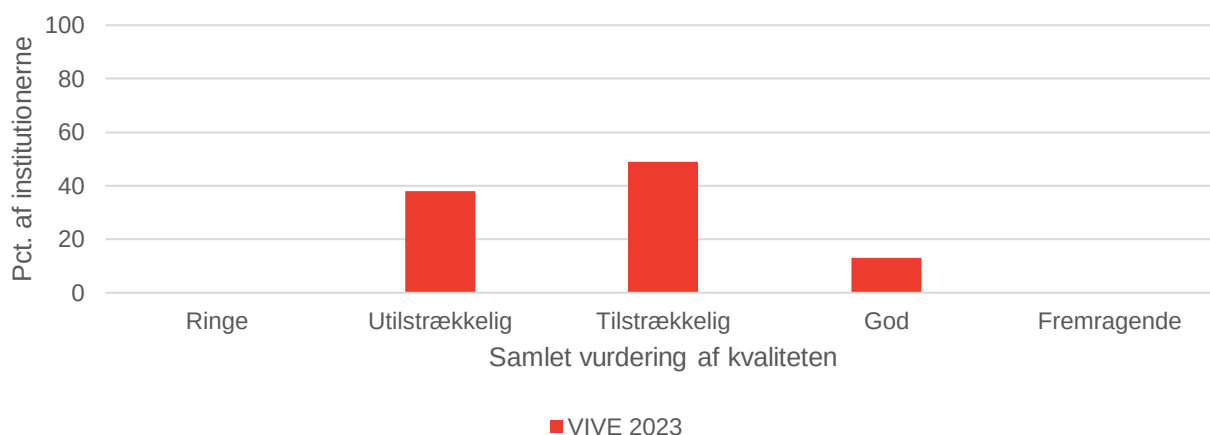
Figur 6 // Resultatet EVAs nationale ECERS-undersøgelse af kvalitet i kommunale børnehaver (EVA, 2020)



Kilde: EVA (EVA, 2020).

Anm.: Se kilde for detaljeret beskrivelse. Kvaliteten er vurderet på baggrund af en analyse med værktøjet ECERS-3 og fordelt over de fire kvalitetskategorier ovenfor.

Figur 7 // Resultater fra EVA og VIVEs nationale KIDS-undersøgelse af kvalitet i pædagogiske læringsmiljøer i dagplejer og vuggestuer (EVA/VIVE, 2023)



Kilde: VIVE og EVA (EVA & VIVE, 2023).

Anm.: Se kilde for detaljeret beskrivelse. Kvaliteten er vurderet på baggrund af KIDS. Der er færre end tre stuer, der vurderes til "ringe" kvalitet. Ingen stuer vurderes til at have "fremragende" kvalitet.

Sammenholder man resultaterne af kommunerens egne vurderinger med EVA og EVA/VIVEs vurderinger, peger tendenserne i markant forskellige retninger. Kommunernes egne KIDS-vurderinger finder varierende kvalitet, der dog overordnet tenderer mod det meget positive. EVA og EVA/VIVEs vurderinger finder ligeledes varierende kvalitet, men de overordnede tendenser peger mod en mere negativ vurdering.

På trods af ovennævnte forbehold for direkte sammenligning giver disse forskelle anledning til spørgsmål om sammenhængen mellem målt og reel kvalitet, samt hvorvidt resultater fra anvendelsen af redskaber, som fx KIDS, afhænger af, hvem der bruger dem. Det er da også et velkendt fænomen, både i dansk og international kontekst, at interne vurderinger generelt ligger højere end eksterne vurderinger (Samuelsson & Sheridan, 2009). Dette skyldes til dels, at interne observatører ikke altid formår at skelne mellem den reelt observerede kvalitet og de pædagogiske intentioner bag den observerede praksis. Skaberne af KIDS anbefaler derfor, at interne KIDS-scorer suppleres med eksterne for at sikre en fyldestgørende vurdering af kvaliteten (Ringsmose & Kragh-Müller, 2020).

05

Sprog værdi som udtryk for kvalitet

Sprogværði som udtryk for kvalitet

Dette kapitel omhandler vores tilgang til de tilsynsrapporter, hvor kvalitetsvurderingen primært er struktureret sprogligt. Det er de rapporter, vi her kalder 'narrative'. Dvs. tilsynsrapporter, der ikke bruger standardiserede, numeriske mål for kvalitet, som fx KIDS og ECERS gør.

Eftersom mængden af narrative tilsynsrapporter er forholdsvis stor, har vi udviklet et specialværktøj, der ved hjælp af en såkaldt sprogværði kvantificerer, hvor positivt eller negativt rapporternes sprog er. Sprogværdien er baseret på formodningen om, at en rapport, der beskriver høj kvalitet, vil være skrevet i et mere positivt sprog end en rapport, der beskriver lav kvalitet. Det er denne ladning i sproget, vi her forsøger at sætte et tal på i form af sprogværdien. For detaljer om denne metode henviser vi til metodebilaget bagest i denne rapport.

BOKS 1: Eksempler på sprogværdier

Når vi anvender værktøjet på den samlede mængde af rapporter, får vi en samlet sprogværði for hver rapport på sætningsniveau. Som eksempel på, hvordan det fungerer, tager vi udgangspunkt i en tilsynsrapport for en vuggestue.

Denne tilsynsrapport indeholder formuleringer som "vuggestuen bliver underlagt skærpet tilsyn" og "bekymringerne bygger på, at institutionen har været påvirket af massiv personaleomsætning, lederskift og fald i børnetal". Ordvalget i disse sætninger bevirker, at rapportens overordnede sprogværði trækkes ned. Desuden har rapporten formuleringer som "børnene har gode muligheder for aktiv og meningsgivende deltagelse", og "fortsæt den gode praksis", der trækker den samlede sprogværði op. Rapporten beskriver også, at "forældrebestyrelsen vil blive inddraget i implementering af handleplanen", en sætning der i vores model tildeles værdien 0.

Den samlede sprogværði for denne vuggestues tilsynsrapport er blandt de absolut laveste i vores datasæt. Dette reflekterer umiddelbart også indtrykket af rapporten som helhed, hvilket også bekræftes af vores robusthedstjek af vores værktøj (se metodebilaget for detaljer).

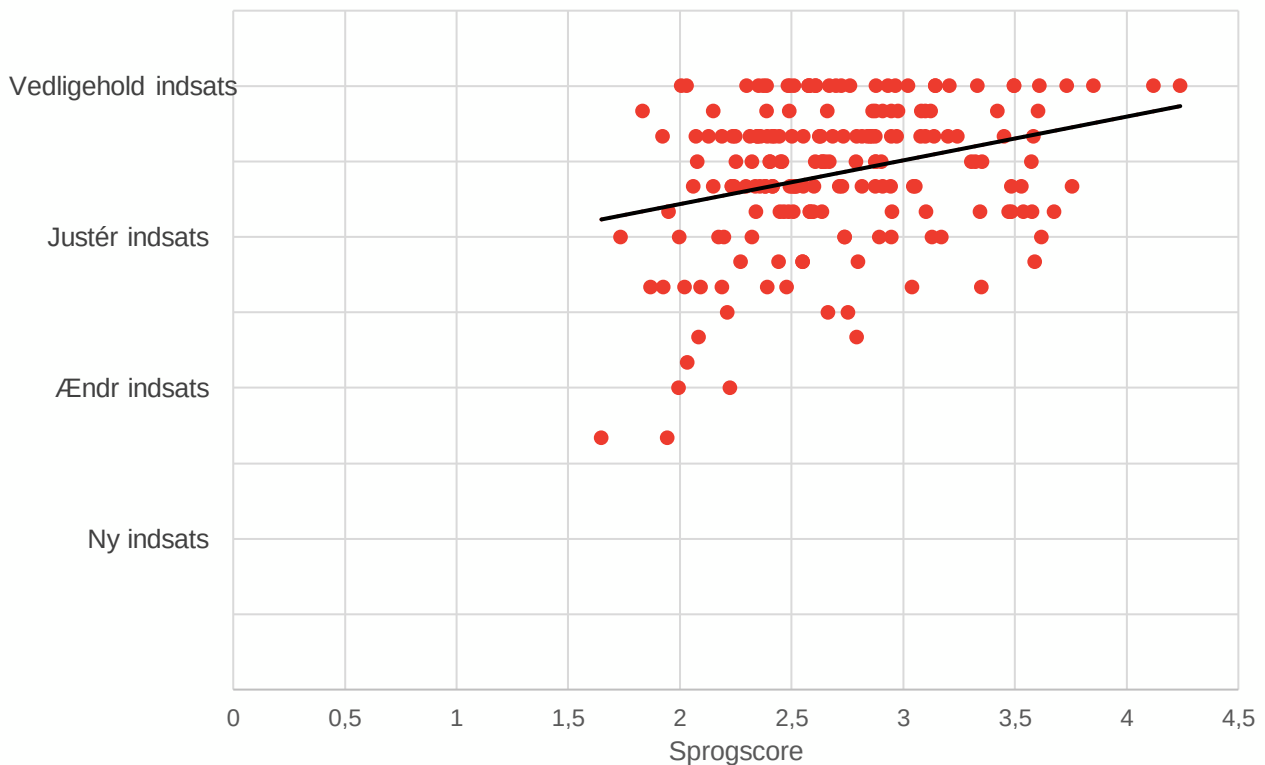
Sprogværði i forhold til kvalitetsvurderinger

For yderligere at teste vores sprogværði op mod kvalitetsvurderinger kan vi se på rapporter, der indeholder både en kvantitativ score og en skriftlig vurdering af kvaliteten. Et sådant eksempel finder vi i Københavns Kommune. Her fokuserer tilsynet på seks temaer, der hver især får en vurdering på en ordinal skala, der går fra "ny indsats" til "vedligehold indsats". "Ny indsats" dækker over, at dagtilbuddet skal ændre deres praksis væsentligt, hvorimod "vedligehold indsats" indikerer, at institutionen gør det godt og blot skal fortsætte (Københavns Kommune, 2023).

Vi har omregnet den ordinale skala til talværdier fra 1-4 og taget gennemsnittet på tværs af de seks temaer. Som det ses i figur 8, er der en tydelig sammenhæng mellem sprogværdien og den overordnede vurdering. Således kan sprogværdien alene forklare 9 pct. af variationen i gennemsnittet fra den ordinale skala. Tilsvarende er sprogværdien statistisk signifikant, hvad angår sammenhæng mellem KIDS-scorer og sprogværdier

i de kommuner, der har både KIDS-scorer og kommentarer. For yderligere detaljer henviser vi til metodebillet.

Figur 8 // Sammenhæng mellem sprogværdien og vurderingen af indsatsen i daginstitutioner i København



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioden 1.5.2022 til 2024

Anm.: Hver prik repræsenterer en af n = 188 tilfældigt udvalgte tilsynsrapporter fra Københavns Kommune, ekskl. rapporter uden vurdering. Forklaringsgraden for den lineære trend er 9 pct., og sammenhængen er signifikant på et 0,1 pct. signifikansniveau.

Sprogværdis som udtryk for andet end kvalitet

Hvis sprogværdien skal kunne læses som et udtryk for den vurderede kvalitet, kræver det, at der ikke er andre faktorer, der kan forklare forskellene i sprogværdis. Vi har her overvejet to forhold, der kunne tænkes at påvirke sprogværdien, nemlig sprogbrug og sætningslængde.

I forhold til sprogbrug kræver vores tilgang, at der ikke er væsentlige forskelle med hensyn til, hvordan rapporternes forfattere udtrykker sig. Hvis én tilsynsførende fx udtrykker sig i overstrømmende positive vendinger, uafhængigt af hvad vedkommende observerer, mens en anden udtrykker sig mere pessimistisk, ville det give sig udslag i sprogværdierne. Vi har ikke en direkte test for, om sprogbrugen er ens på tværs af rapporter. Men rapporterne er i overvejende grad udarbejdet af fagpersoner med en baggrund i pædagogik, ligesom de omhandler de samme tematikker og følger samme struktur inden for hver kommune. Samtidig ved vi fra tidligere undersøgelser, at der lader til at være større variation inden for de enkelte kommuner, end der er mellem kommunerne (DEA, 2020, 2021).

06

Narrative rapporteur

Narrative rapporter

I dette kapitel ser vi nærmere på indholdet af de rapporter, vi kalder 'narrative', og som udgør langt hovedparten af begge vores datasæt. Denne betegnelse dækker over forskellige rapport-strukturer, hvor nogle er rent tekstbaserede, mens andre kombinerer tekst og tal.⁶ Langt de fleste af disse rapporter er struktureret omkring en række tematikker eller kriterier, der enten kan vurderes efter prædefinerede kategorier eller i en mere fortællende form. Der er således tale om mange forskellige typer af rapporter, der dog har det til fælles, at tilsynets vurdering af kvalitet primært formidles sprogligt og derfor her behandles separat fra KIDS-rapporterne. Den nærmere tilgang er beskrevet i forrige kapitel samt i metodebilaget.

Sprogværdier over tid

Mellem 1. og 2. datasæt bliver rapporterne i gennemsnit både længere og mere positive, som det fremgår af i tabel 3. Da vi kun har to datapunkter i form af vores to datasæt, kan vi ikke her afgøre, om ændringen er en del af en generel tendens, eller om den specifikt skyldes lovændringen om styrket tilsyn.

Tabel 3 // Spredning i sprogværdier

	1. datasæt	2. datasæt	
Alle tilsynsrapporter	<i>n</i> = 2.122	<i>n</i> = 2.877	
Gennemsnitlig sprogværdi	2,250	2,289	*
Gennemsnitligt antal sætninger	145,574	159,649	***
Spredning i sprogværdi (std.-dev.)	0,857	0,714	***
Tilsynsrapporter i begge datasæt	<i>n</i> = 1.409	<i>n</i> = 1.409	
Gennemsnitlig sprogværdi	2,284	2,367	***
Gennemsnitligt antal sætninger	142,341	160,405	***
Spredning i sprogværdi (std.-dev.)	0,878	0,757	***

Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024

Anm.: *** *p* < 0,01, ** *p* < 0,05, * *p* < 0,1. Rapporter med en gennemsnitlig sprogværdi på over 7 er fjernet her, og det samme gælder rapporter med en længde på over 500 sætninger.

Som nævnt i forrige kapitel kan vi heller ikke direkte sammenligne niveauet for sprogværdier mellem kommuner, der benytter forskellige skabeloner. Det betyder bl.a., at vi ikke entydigt kan lægge os fast på en betydning af, at den gennemsnitlige sprogværdi er steget fra 1. datasæt til 2. datasæt.⁷ Det er dog værd at bemærke, at den gennemsnitlige sprogværdi stiger for de institutioner, vi kan følge over tid, jf. den nederste halvdel af tabel 3. Hvis tilsynene er blevet mere systematiske som et resultat af lovændringen, ville vi forvente, at der både bliver lavet flere positive, men i særdeleshed flere negative beskrivelser af kvaliteten i institutionerne. Det ser vi dog ingen tegn på her, hvilket kunne tyde på, at der stadig er visse uoverensstemmelser mellem reel og vurderet kvalitet.

⁶ De narrative rapporter indbefatter således også Københavns Kommunes tilsynsrapporter og rapporter udarbejdet efter Rambølls "Hjernen og hjertet"-koncept (Rambøll, 2024).

⁷ Teknisk set kan skabelonteksten også påvirke spredningen i scorer, da fx rapporter med en meget kort skabelontekst vil have mere volatile sprogscore end rapporter, hvor skabelonteksten udgør en stor andel af den samlede tekstmængde. Vi afstår derfor fra at konkludere på spredningen af sprogscore for de rapporter, hvor skabelonteksten udgør størstedelen af teksten.

I forhold til spredningen i sprogverdier er der sket et fald i spredningen fra 1. datasæt til 2. datasæt. Det gælder, både når vi kigger på alle tilsynsrapporter, og når vi fokuserer på de institutioner, vi kan følge over tid, jf. tabel 3. Det betyder, at sprogverdierne ligger tættere på hinanden nu end tidligere. Den mindre spredning i sprogverdier kan skyldes flere faktorer:

1. **Det afspejler en reel mindre spredning i kvalitet.** Der var med andre ord flere almindeligt gode institutioner i 2. datasæt end i 1. datasæt, hvor der til gengæld var flere institutioner med meget lav og meget høj kvalitet.
2. **De nye tilsynsrammer har medført øget systematik.** Når alle kommuner er forpligtet til at føre regelmæssige tilsyn, så stiger antallet af tilsyn. Det betyder, at tilsynene sættes i system, og det kan inkludere et øget fokus på metodisk stringens.
3. **Tilsynene gennemføres i flere gennemsnitlige institutioner.** Tidligere kan tilsynene have haft et (evt. ubevidst) fokus på institutioner med meget høj eller meget lav kvalitet, hvilket tilsynsrapporterne vil afspejle. Når regelmæssige tilsyn gennemføres i alle institutioner, vil der efter lovændringen være flere tilsyn i på tværs af alle institutioner.
4. Endelig kan der være tale om en kombination af de ovennævnte faktorer.

Hvad den præcise forklaring på den mindre spredning er, er det vanskeligt at fastslå ud fra de her anvendte data. Men forklaringen har indflydelse på, om man bedst adresserer forskellene ved at sætte ind over for kvaliteten i institutionerne, kvaliteten i tilsynet eller begge dele.

Sammenhæng i sprogverdier over tid

Forrige kapitel beskriver, hvordan institutioner, der i 1. datasæt har fået en høj KIDS-score, med stor sandsynlighed også får det i 2. datasæt. Den tendens går umiddelbart igen i de narrative rapporter, hvor vi har testet forskellige modeller for sammenhængen mellem sprogverdier i 1. og 2. datasæt, jf. tabel 4.

I tabel 4 viser model 1, at sprogverdien i 1. datasæt overordnet set hænger sammen med sprogverdien i 2. datasæt.⁸ Hvis nogle kommuner konsekvent over hele perioden har en høj sprogverdi som følge af deres skabelon, kan det dog give en kunstigt stærk sammenhæng. For at tage højde for det, viser model 2, at kommuneforskelle kun forklarer en del af sammenhængen over tid. Model 3 inkluderer en interaktion mellem sprogverdien i 1. datasæt og kommuneeffekterne, dvs. en kombination af model 1 og model 2.

I model 3 er sprogverdien i 1. datasæt statistisk signifikant med hensyn til forklaringen af sprogverdien i 2. datasæt for nogle specifikke kommuner (se bilag for detaljer). Det betyder, at der er en tendens til, at institutioner, der vurderes fx meget negativt i første periode, også vurderes negativt i anden periode, men at sammenhængen er drevet af nogle specifikke kommuner.

⁸ Vi har her kun set på kommuner, hvor der er rapporter for flere end ti institutioner, som vi kan følge i begge perioder. Det reducerer datasættet til 40 kommuner og 1.245 observationer.

Tabel 4 // Sammenhæng i sprogverdier over tid

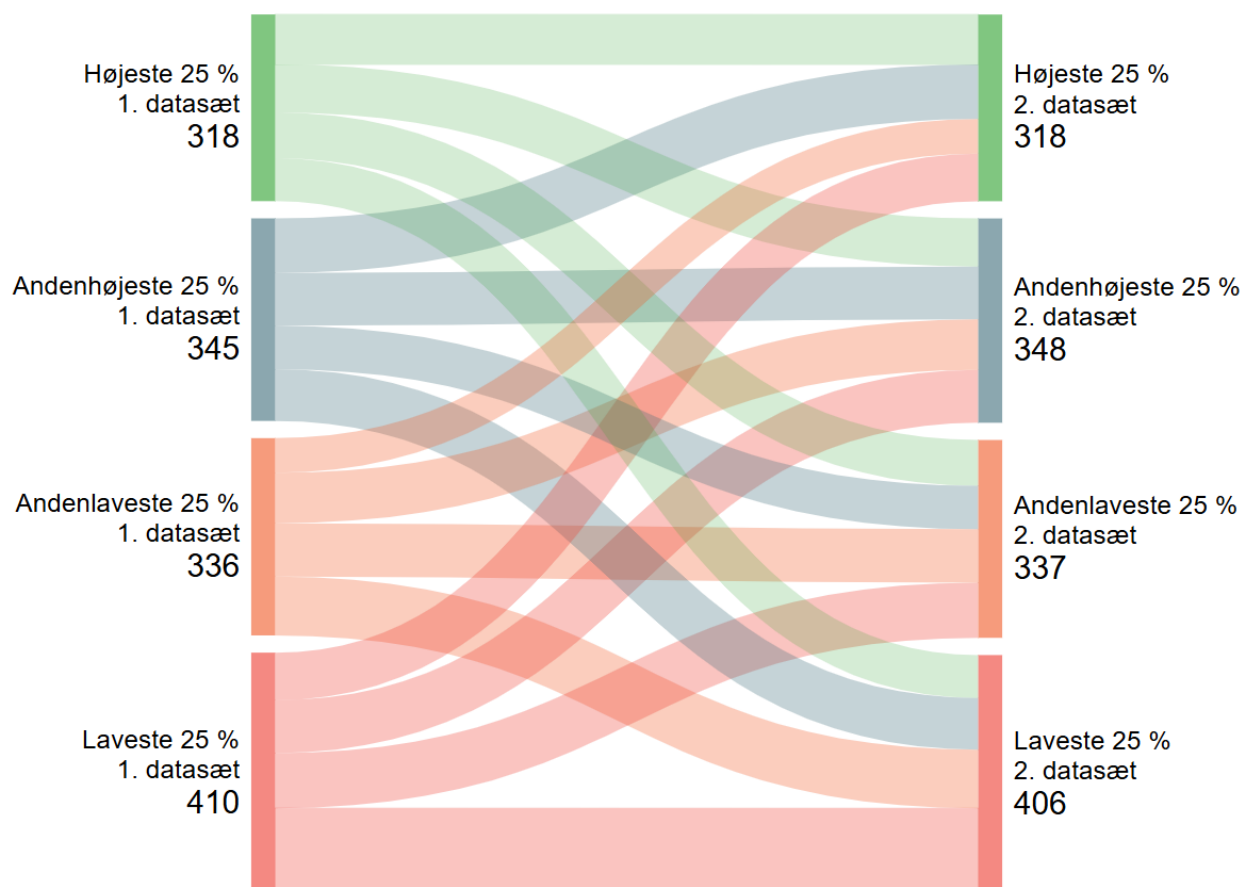
	Model 1	Model 2	Model 3
Input:	Sprogverdi i 1. periode	Sprogverdi i 1. periode og kommuneeffekter (FE)	Sprogverdi i 1. periode, kommuneeffekter (FE) og interaktion mellem kommuneeffekter og sprogverdi i 1. periode
Sprogverdi, 1. periode	0,226*** (0,030)	0,143*** (0,027)	. .
Antal observationer	1.242	1.242	1.242

Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024

*Anm.: n = 1.242 fordelt på 40 kommuner. *** p < 0,01, ** p < 0,05, * p < 0,1. Robuste standardfejl i parentes. Rapporter med en gennemsnitlig sprogverdi på over 7 er fjernet her, og det samme gælder rapporter med en længde på over 500 sætninger. Rapporter fra kommuner, hvor vi kan følge 10 eller færre institutioner over tid er ikke inkluderet. For en illustration af fordelingen af sprogverdierne i hhv. 1. og 2. datasæt se bilag 5.*

Figur 9 viser sammenhængen i sprogverdier over tid, dvs. fra 1. datasæt til 2. datasæt, fordelt på kvartiler. Dvs. at de 25 pct. af tilsynsrapporterne med de højeste sprogverdier ligger i 4. kvartil, mens dem med de laveste sprogverdier ligger i 1. kvartil inden for hver kommune. Som det ses i figuren, varierer sammenhængen i sprogverdier på tværs af kvartilerne. Mens der for rapporter med de højeste, de næsthøjeste og de næstlaveste sprogverdier er en vis grad af mobilitet, har de rapporter i 1. kvartil med de laveste sprogverdier i betydeligt større sandsynlighed for også at få en værdi blandt de laveste i 2. datasæt. Faktisk har 35 pct. af tilsynsrapporterne i det laveste kvartil også de laveste sprogverdier i 2. datasæt. Det er dermed hovedsageligt for de dårligst vurderede daginstitutioner, at der er en stærk sammenhæng i sprogverdierne over tid.

Figur 9 // Relativ vurdering af kvaliteten i daginstitutionerne over tid



*Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024
Anm.: n = 1.409. Rapporter med en gennemsnitlig sprog værdi på over 7 er fjernet her, og det samme gælder rapporter med en længde på over 500 sætninger. Institutioner, der ikke optræder i begge perioder, er ikke vist.*

Mobilitet mellem kvartilerne er ikke nødvendigvis overraskende, i lyset af at de nye tilsynsrammer kan have haft en effekt på, hvordan institutionerne vurderes. Dette står i kontrast til tidligere analyser, der peger på en høj grad af kontinuitet med hensyn til institutionernes resultater (DEA, 2021; 2024). En forklaring kan være, at kvaliteten i institutionerne varierer markant mere, end tidligere undersøgelser har fundet. En anden forklaring er dog, at ændringer i kommunernes tilsynspraksis giver andre resultater, selvom kvaliteten reelt er den samme. Det vil betyde, at tilsynsrapporternes brugsværdi for fx forældre er stærkt begrænset.

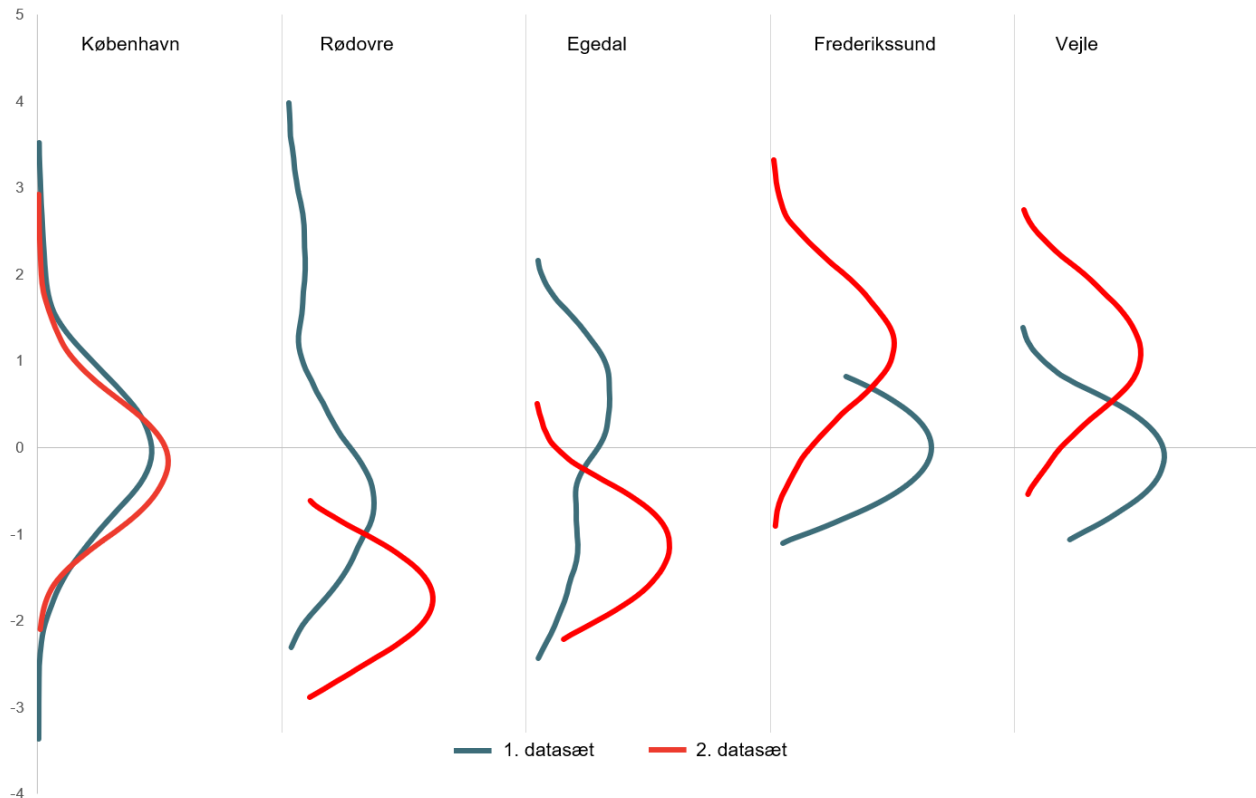
Det er til gengæld værd at bide mærke i den forholdsvis store gruppe af institutioner, der vedvarende får en mere kritisk vurdering af kvaliteten. Dette kan være et tegn på, at der er en relativt konstant gruppe daginstitutioner, der ikke får den støtte og opfølgning, de har brug for for at kunne forbedre deres kvalitetsvurdering.

Kommunale forskelle i sprog værdispredning

På samme måde kan vi også undersøge spredningen i sprog værdi på kommuneniveau. Mens vi generelt er forbeholdne over for at sammenligne kommunernes sprog værdier, er der bedre belæg for at sammenligne sprog værdispredninger. På den måde kan vi ved at zoome ind på forskellige kommuner og se på udviklingen

i spredningen i sprogværdis, sammenholdt med deres specifikke tilsynspraksis, forsøge at indkredse de fire mulige forklaringsmodeller i forhold til udviklingen i spredning, som vi opridsede på side 26. Figur 10 illustrerer, hvordan spredningen i sprogværdis har udviklet sig fra 1. datasæt til 2. datasæt for fem udvalgte kommuner.⁹ Spredningerne i figuren er centreret omkring den gennemsnitlige sprogværdis for kommunen i 1. datasæt. Når spredningen for rapporterne i 2. datasæt skubbes hhv. op eller ned, illustrerer det, om kommunens sprogværdier i gennemsnit er forbedret eller forværret.

Figur 10 // Spredning i sprogværdier for udvalgte kommuner



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024
Anm.: Der indgår i hhv. 1. og 2. datasæt for hver kommune vist $n = 276$ og $n = 368$, $n = 20$ og $n = 20$, $n = 20$ og $n = 26$, $n = 39$ og $n = 36$, $n = 24$ og $n = 58$ (antal rapporter). Rapporter med en gennemsnitlig sprogværdis på over 7 er fjernet her, og det samme gælder rapporter med en længde på over 500 sætninger. Spredningen ('kernel density') er udregnet for alle rapporter, ekskl. ovenstående, i hver periode. Spredningen er langs x-aksen defineret i rummet $[0;1[$, for hver kommune.

Som det fremgår af figur 10, er der stor forskel på, hvor meget spredningen ændrer sig fra 1. datasæt til 2. datasæt. I den ene ende af skalaen ligger Københavns Kommune, hvor spredningen er blevet reduceret (dvs. at kurvens top- og bundpunkter ligger tættere på hinanden), samtidig med at mønsteret bibeholdes (dvs. at kurvens midtpunkt og hældning ikke ændrer sig markant). Dette skal muligvis ses i sammenhæng med, at Københavns Kommune efter lovændringen har beholdt sin tilsynsmodel, som er blevet udviklet over en længere periode og i forvejen levede op til de styrkede tilsynskrav (Københavns Kommune, 2023). Dette kunne underbygge forklaringsmodel 2 ovenfor, hvor et mere metodisk og konsistent tilsyn kan medføre

⁹ Fordelingen for alle kommuner er vist i bilag 4.

mindre spredning i sprogverdier. En sådan systematisk tilgang ville desuden give bedre mulighed for både at sætte målet ind over for kvalitetsudfordringer og at spore kvalitetsændringer over tid

For de fire kommuner i højre side af figur 10 gælder det, at spredningen i sprogverdi varierer betydeligt over en relativt kort tidsperiode. Vi ser højere gennemsnitligt vurderet kvalitet og en større spredning i vurderet kvalitet i rapporterne fra Vejle og Frederikssund, og modsvarende er rapporterne fra fx Rødovre og Egedal blevet mere konsistente i deres (mere negative) vurdering af kvaliteten. Nogle af disse kommuner, som fx Rødovre, lader til at have gjort deres rapportformat mere detaljeret, så det nu fx inkluderer observationer fra tilsynet. Formentlig vil en sådan ændring i praksis kunne give et mere retvisende billede af kvaliteten i institutionerne, men det betyder også, at det bliver vanskeligt at spore kvalitetsændringer over tid.

Tilsammen peger disse resultater på to forhold. For det første giver de forskellige kommunale tilsynspraksisser forskellige spredninger i sprogverdi og dermed i vurderingen af kvalitet. For det andet giver et skift i tilsynspraksis udsving i forhold til sprogverdi, hvilket kan vanskeliggøre en konsistent vurdering af kvalitet. Igen gør disse udsving det vanskeligt for læserne af tilsynsrapporter at danne sig et overblik over den reelle kvalitet i en kommunes daginstitutioner og træffe beslutninger derefter.

07

Litteraturliste

Litteraturliste

- Bekendtgørelse af lov om dag-, fritids- og klubtilbud m.v. til børn og unge, LBK nr 824 af 15/08/2019 (2019). <https://www.retsinformation.dk/eli/lt/2019/824>
- BUVM. (2020). *Styrket tilsyn | Aftale om minimumsnormeringer*. <https://www.uvm.dk/dagtilbud/love-og-regler-formaal-og-aftaler/aftale-om-minimumsnormeringer/styrket-tilsyn>
- DEA. (2020). *Omgængere i børnehaveklassen*. <https://dea.nu/publikationer/omgaengere-er-udfordret-fagligt-og-socialt>
- DEA. (2021). *Løfteevne i daginstitutioner*. Tænketanken DEA. <https://www.datocms-assets.com/22590/1611930195-lofteevne-i-daginstitutioner-januar-2021.pdf>
- DEA. (2023a). *Kvalitet i dagtilbud ifølge tilsynsrapporter*. <https://dea.nu/publikationer/kvalitet-i-dagtilbud-ifolge-tilsynsrapporter>
- DEA. (2023b). *Trivsel og sygefravær i daginstitutioner—Medarbejdertilfredshed i Aarhus*. <https://www.datocms-assets.com/22590/1704445070-trivsel-og-sygefravaer.pdf>
- DEA. (2024). *Daginstitutionens betydning for børns skolegang*. https://www.datocms-assets.com/22590/1712741921-daginstitutionens-betydning-for-borns-skolegang_final.pdf
- DSL. (2024). *Danish Sentiment Lexicon*. https://sprogteknologi.dk/da_DK/dataset/14a3e07c-50e9-42b8-b06c-dd965d203f55
- EVA. (2015). *Pædagogisk tilsyn*. <https://www.eva.dk/dagtilbud-boern/paedagogisk-tilsyn>
- EVA. (2020). *Læringsmiljøkvalitet—National undersøgelse*. Danmarks Evalueringsinstitut. <https://www.eva.dk/dagtilbud-boern/laeringsmiljoekvalitet-national-undersoegelse>
- EVA. (2022). *Erfaringer med at undersøge kvalitet nationalt*. <https://eva.dk/Media/638337457417877727/Erfaringer%20med%20at%20unders%C3%B8ge%20kvalitet-Forarbejde%20til%20national%20kvalitetsunders%C3%B8gelse%20L3.pdf>
- EVA, & VIVE. (2023). *Kvalitet i dagtilbud: National undersøgelse af kvalitet i pædagogiske læringsmiljøer og rammer i kommunale daginstitutioner og dagplejen for 0-2-årige børn*. VIVE.
- FOA, & Bureau 2000. (2020). *Tilsyn med dagpleje og daginstitutioner 2020*. https://www.bureau2000.dk/CustomerData/Files/Folders/7-seneste-nyheder/1086_tilsyn-med-dagpleje-og-daginstitutioner-2020.pdf
- Harms, T., Clifford, R. M., & Cryer, D. (2014). *Early Childhood Environment Rating Scale, third edition (ECERS-3)*. Teachers College Press.
- Københavns Kommune. (2023). *Rammer for pædagogisk tilsyn i dagtilbud*. <https://www.kk.dk/sites/default/files/2023-05/TIL%20KK.dk%20-%20Rammer%20for%20tilsyn%20i%20dagtilbud.pdf>
- Lauridsen, G. A., Dalsgaard, J. A., & Svendsen, L. K. B. (2019). SENTIDA: A New Tool for Sentiment Analysis in Danish. *Journal of Language Works - Sprogvidenskabeligt Studentertidsskrift*, 4(1), 38–53.
- Rambøll. (2024). *Hjernen og hjertet*. <https://hjernenhjertet-dk.ramboll.com/>
- Ringsmose, C., & Kragh-Müller, G. (2020). *KIDS – Kvalitetsudvikling i daginstitutioner* (2. udg.). Dansk Psykologisk Forlag.
- Samuelsson, I. P., & Sheridan, S. (2009). Preschool Quality and Young Children's Learning in Sweden. *International Journal of Child Care and Education Policy*, 3(1), 1–11.
- Schneidermann, N. S., & Pedersen, B. S. (2022). Evaluating a New Danish Sentiment Resource: The Danish Sentiment Lexicon, DSL. *Proceedings of the SALLD-2 Workshop*, 19–24. <https://aclanthology.org/2022.salld-1.4.pdf>

08

Bilag

Bilag

Bilag 1. Data- og metodebeskrivelse

Om data og dataindsamling

Daginstitutioner i alle landets kommuner er identificeret via STIL's dagtilbudsregister. På baggrund af institutionens navn og placering søger vi efter offentligt tilgængelige tilsynsrapporter på kommunens og daginstitutionens hjemmeside. Det er gjort i forbindelse med første indsamlingsperiode (DEA, 2023a) og igen i anden periode. De to dataindsamlingsprocesser har resulteret i hhv. et 1. datasæt og et 2. datasæt over tilsynsrapporter, der dækker perioderne 2018 til 30.04.2022 og 01.05.2022 til 2024. For det seneste datasæt er institutionstype fra dagtilbudsregistret desuden koblet til tilsynsrapporter for den pågældende institution. Vi kan følge udviklingen i tilsynsrapporterne for en institution over tid ved at sammenholde rapporterne i 2. datasæt med rapporterne i 1. datasæt. Se kapitlet "Overblik over tilsynsrapporter i denne analyse" og notatet fra første indsamlingsperiode (DEA, 2023a) for en detaljeret gennemgang af datasættenes størrelse og indhold. 104 ens rapporter findes i begge datasæt. Vi har ens rapporter i de to datasæt, fordi nogle rapporter er tomme, og fordi vi har hentet enkelte rapporter i 2. datasæt, der allerede var offentliggjort i den første periode, når det fx var offentliggjort som "seneste" tilsyn. Disse dubletter er ikke medregnet i analyserne.

I begge datasæt undlader vi at bruge tilsynsrapporter med flere end 500 sætninger, da de ikke er repræsentative for den samlede mængde af rapporter. Det skyldes, at nogle kommuner offentliggør tilsynsrapporter på fx klyngeniveau, hvilket ikke harmonerer med vores design, hvor vi fokuserer på udviklingen i de enkelte institutioner. Derudover udelader vi rapporter med ekstreme sprogverdier, dvs. en gennemsnitlig sprogverdi på over 7 eller under 0. Dette er for ikke at overrepræsentere det fåtal af rapporter, der ikke deler form eller sprogbrug med flertallet af rapporter. Det drejer sig om 65 og 19 tilsynsrapporter i hhv. første og anden dataindsamling.

Sprogverdierne er baseret på en såkaldt *sentiment analysis* på baggrund af Danish Sentiment Lexicon fra Det Danske Sprog- og Litteraturselskab (DSL) og Center for Sprogteknologi ved Københavns Universitet (DSL, 2024). Her matcher vi ord for ord leksikonets 14.000 ord, der alle er tildelt en score i spændet $[-5; 5]$, med tekstindholdet i tilsynsrapporterne. Scoren aggregerer vi over hele sætninger, og vi definerer sprogverdien som den gennemsnitlige score over antallet af sætninger i en given rapport. Det betyder, at sprogverdien er følsom over for fx en udvikling i sætningslængden. Den gennemsnitlige sætningslængde er dog konstant på 19 ord pr. sætning i begge datasæt. Sprogverdien er dermed kontinuert defineret i rummet $] - \infty; \infty[$.

DSL fanger ikke den begrebssammenhæng, ordene indgår i, og fx styrkemarkører såsom "meget" tildeles ikke en vægt. Implikationerne af dette på scorer forventes at være neutrale, da effekten er den samme for både positive og negative ord. Ord uden ladning tildeles scoren 0. Metoden forudsætter, at vi kan ekstrahere tekst fra rapporterne. Rapporter offentliggjort som fx scan udgår derfor, og fx diagrammer eller tabeller indgår kun i sprogverdien i det omfang at vi kan aflæse og score deres tekstindhold.

Regnemetoder

For at sammenligne KIDS-scorer, der er defineret i forskellige spænd, udregner vi en vægtet KIDS-score, der er defineret relativt til det spænd, scoringen foregår i. På rapportniveau definerer vi $S_w = (S - L)/(H - L)$, hvor S er den samlede score, L og H er den hhv. laveste og højeste mulige score. S_w kan sammenlignes på tværs af institutioner, hvor L og H ikke er ens. L, H og S afhænger af antallet af spørgsmål, der indgår i scoren, mens

S_w er renset for den effekt. Vi fortolker i overensstemmelse med EVA og VIVEs metode fra den nationale kvalitetsundersøgelse (EVA & VIVE, 2023) den vægtede KIDS-score som udtryk for kvaliteten i kategorierne Ringe (0-12,5 pct.), Utilstrækkelig (12,5-37,5 pct.), Tilstrækkelig (37,5-62,5 pct.), God (62,5-87,5 pct.) og Fremragende (87,5-100 pct.).

Et enkelt sted normaliserer vi sprogverdierne omkring den gennemsnitlige sprogverdi for rapporterne i kommunen. Dermed fjerner vi niveauforskelle i sprogverdier for forskellige kommuner ved at parallelforskyde sprogverdier op eller ned omkring den gennemsnitlige sprogverdi. Den normaliserede sprogverdi er udregnet som: $\tilde{x}_{i,j,M} = x_{i,j,M} - \bar{X}_{j,M}$, hvor $x_{i,j,M}$ er sprogverdien for rapport i fra kommune j i datasættet $M \in (1; 2)$. $\bar{X}_{j,M}$ er den gennemsnitlige sprogverdi for rapporterne i kommune j i datasættet M . Vi normaliserer desuden sprogverdier i 2. datasæt omkring kommunegennemsnittet i 1. datasæt: $\tilde{x}_{i,j,M=2} = x_{i,j,M=2} - \bar{X}_{j,M=1}$.

Vi anvender en almindelig lineær regressionsmodel til at påvise korrelationer mellem fx længde og sprogverdi af rapporter. Den er defineret som: $y = \beta_0 + \beta_1 x + \varepsilon$. Når vi refererer til, hvor stor en del af variansen i y som x beskriver, så menes der forklaringsgraden, R^2 . Vi udvider ovenstående regressionsmodel til at indeholde kommuneeffekter (*fixed effects*, FE) og interaktionsled i undersøgelsen af persistens i sprogverdier. Dermed omskrives modellen til $y_{\text{sprogverdi},M=2} = \beta_0 + \beta_1 x_{\text{sprogverdi},M=1} + \beta_2 (x_{\text{sprogverdi},M=1} \cdot Z) + \gamma Z + \varepsilon$, hvor β_1 og β_2 er parameterestimatet hhv. for sprogverdien og for interaktionsleddet mellem sprogverdien og kommunerne. Z er kommunerne og γ er kommuneeffekten (FE).

For at illustrere strømme fra kvartiler i 1. datasæt til 2. datasæt, så udregner vi kommunespecifikke kvartilsgrænser for begge datasæt og opdeler alle rapporter, vi kan følge over tid, inden for kvartilerne. Dernæst aggregerer vi over alle kommuner, hvilket dog mekanisk indebærer en ubalance i antallet af observationer inden for hver kvartil.

Begrænsninger i brugen af data

Vi sammenligner KIDS-scorer på tværs af kommuner, selvom der er forskellige tilsynsførende i forskellige kommuner. Brugen af samme værktøj er derfor ikke en garanti for, at kvaliteten bliver vurderet ens på tværs af kommuner. Vi antager, at KIDS-værktøjet og den medfølgende manual giver en tilpas grundig indføring i brugen af værktøjet til, at vi rimeligt kan sammenligne på tværs af kommuner. Den antagelse er underbygget af, at alle kommuner når frem til næsten de samme gennemsnitlige KIDS-scorer, jf. [tabel 3](#). Med det in mente kommenterer vi dog ikke på niveauet af KIDS-scorerne mellem kommunerne.

Ligeledes er der en naturlig begrænsning for sammenligning af sprogverdier. Sprogverdierne, og spredningen i sprogverdierne, kan ud over at afspejle den vurderede kvalitet afhænge af rapportens sætningslængde, hvilken skabelon rapporten er skrevet i, hvem der udfører tilsynet, og hvilken sprogbrug der er anvendt. Som beskrevet ovenfor er sætningslængden i gennemsnit den samme i begge datasæt, og vi ser i data generelt ikke store forskelle her. Vi finder det derfor rimeligt, at forskelle i sprogverdier ikke i betydeligt omfang skal tilskrives forskelle i sætningslængde.

Rapportskabelonen er derimod afgørende for sprogverdierne. Derfor anvender vi ikke sprogvurderingerne til at sige noget om, hvilke kommuner der vurderer kvaliteten i deres institutioner højest/lavest, da niveauforskelle kan tilskrives forskellige rapportskabeloner i de forskellige kommuner. Samtidig er der en stribe kommuner, der anvender forskellige rapportskabeloner inden for hver periode. Derfor undlader vi også at fremhæve, fx hvilke institutioner i en given kommune der opnår den højeste/laveste sprogverdi, da det frem for at afspejle forskelle i vurderet kvalitet potentielt afspejler forskelle i rapportskabeloner. Yderligere er der en del kommuner,

der ændrer deres rapportskabelon(er) fra 1. datasæt til 2. datasæt. Alligevel sammenligner vi både niveau og spredning (se figur 10). Det skyldes, at vi her ønsker at illustrere netop den pointe, at det er vanskeligt at sammenligne udviklingen i vurderet kvalitet over tid, når skabelonerne ændres. Vi afstår igen fra at træffe konklusioner på baggrund af niveauet for sprogværddierne.

Vi anvender dog niveauet for sprogværddierne til at undersøge udviklingen i den *relative* vurdering af kvaliteten over tid (se figur 9). Det er muligt, fordi vi relaterer institutionens sprogværddi til sprogværddien for de andre institutioner i kommunen. Når en kommunes skabelon ændres fra datasæt til datasæt, er den relative vurdering derfor upåvirket. Når kommunerne anvender forskellige skabeloner inden for hver periode, så betyder det, at nogle institutioner vurderes relativt bedre/dårligere end andre institutioner i kommunen, uden at det reelt afspejler forskelle i den vurderede kvalitet. Dermed bliver det i højere grad tilfældigt, om en rapport har en høj/lav sprogværddi i forhold til de andre institutioner i kommunen. De) er tilfældet i begge datasæt. I sammenligningen af den relative placering af institutionerne fra 1. datasæt til 2. datasæt, bør mange institutioners relative placering derfor ændre sig. Den faktiske persistens for de institutioner, der har den laveste kvalitet, er med al sandsynlighed større end den persistens, vi kan vise, da en del af de institutioner med dårligst vurderet kvalitet i 1. datasæt får en højere sprogværddi i 2. datasæt på baggrund af ændringer i skabelonen.

Slutteligt er forskelle i, hvem der udfører tilsynet og deres sprogbrug, potentielle kilder til forskelle i sprogværddien. Vi antager, at når det er en relativt homogen faggruppe, der udfører tilsyn, så resulterer deres faglighed i, at forskelle i vurderet kvalitet reelt dækker over forskelle i kvalitet, frem for forskelle i de tilsynsførendes evne til at vurdere kvalitet nøjagtigt. Vi har ikke mulighed for at efterprøve den antagelse. Vi antager desuden, at deres faglighed medfører en domænespecifik sprogbrug, dvs. at de forfatter tilsynsrapporterne nogenlunde ens. Vi ser ikke tegn på, at det ikke skulle være tilfældet, når vi læser tilsynsrapporter fra forskellige kommuner i de to datasæt. Fordi vi ikke konkluderer på niveauforskelle i sprogværddier kommunerne imellem, så tillader vi fx regionale forskelle med hensyn til ladningen i sprogbrugen.

Validitet af sprogværddier

Vi antager i analysen, både at metoden bag vores sprogværddier, *sentiment analysis*, giver en retvisende kvantificering af ladningen i sprogbrugen, og at ladningen hænger sammen med den vurderede kvalitet. I den første antagelse læner vi os op ad Lauridsen et al., der påviser validiteten af et mindre og ældre dansk sentimentleksikon, SENTIDA (Lauridsen et al., 2019), samt Schneidermann og Pedersens (2022) evaluering af det sentimentleksikon, vi har anvendt her (DSL, 2024).

For at påvise en sammenhæng mellem kvalitetsvurderingen og ladningen i sprogbrugen benytter vi os af tilsynsrapporter fra Københavns Kommune. De er både narrative og indeholder en kvantificeret kvalitetsvurdering, som giver os mulighed for at sammenholde den vurderede kvalitet og sprogværddien (Københavns Kommune, 2023). Vi udvælger en stikprøve af rapporter fra Københavns Kommune i begge indsamlingsperioder, omdanner kommunens ordinale kvalitetsskala til en firepunktsskala i intervallet [1;4], og udregner den gennemsnitlige kvalitetsvurdering på tværs af pejlemærker. Sprogværddien forklarer hhv. 12 og 9 pct. af variationen i 1. og 2. datasæt og er statistisk signifikant på et 1-pct.-signifikansniveau i begge datasæt, jf. metodetabel 1. Det er forholdsvis højt, taget i betragtning af, at både sprogværddien og kvalitetsvurderingen er et gennemsnit for hele rapporten, hvilket betyder, at fremtrædende sammenhænge inden for de enkelte pejlemærker er udglattet af gennemsnittet. I KIDS-rapporter med grundige kommentarer, fx fra Slagelse, finder vi samme korrelation, se evt. metodebilag 1.

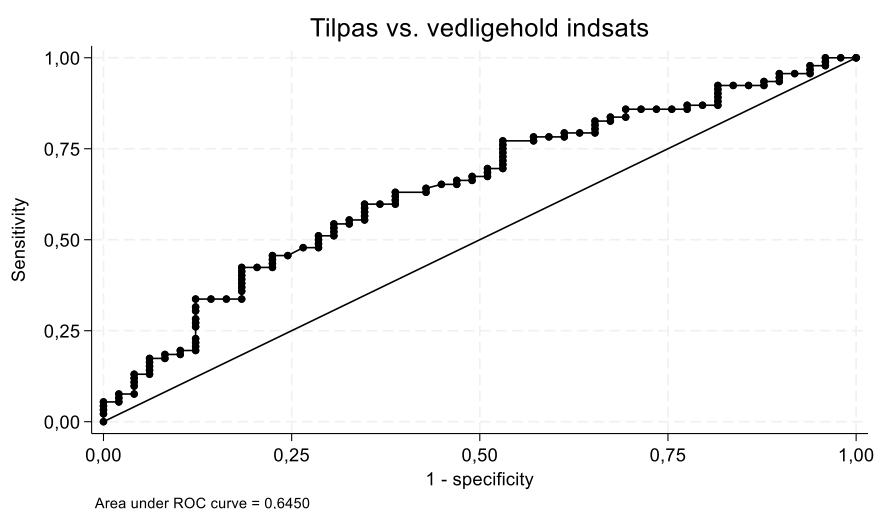
Metodetabel 1 // Validitet af sprogverdier i Københavns Kommune

	1. datasæt	2. datasæt
Sprogværði	0,214 *** (0,041)	0,290 *** (0,080)
Konstantled	1,962 *** (0,135)	2,638 *** (0,226)
R ²	0,117	0,090
Antal observationer	141	188

Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioderne 2018 til 30.4.2022 og 1.5.2022 til 2024
*Anm.: Robuste standardfejl i parentes. *** $p < 0,01$, ** $p < 0,05$, * $p < 0,1$. Stikprøven er repræsentativ for den faktiske fordeling blandt alle rapporter i Københavns Kommune på et 95-pct.-signifikansniveau.*

For at nuancere, i hvilket omfang vores sprogverdier hænger sammen med kvalitetsvurderingerne, kan vi se, i hvor høj grad vi korrekt kan klassificere kommunens kvalitetsvurdering på baggrund af sprogverdierne i det første pejlemærke i 1. datasæt. Her fordeler kvalitetskategorierne sig i kategorierne "tilpas indsats" og "vedligehold indsats". Vi opstiller en model for kvalitetsvurderingen baseret på sprogverdierne (se metodebilag 2), og arealet under ROC-kurven angivet i metodefigur 1 illustrerer, i hvilket omfang modellen kan klassificere kvalitetskategorierne korrekt. Ved en perfekt sammenhæng mellem sprogverdier og kvalitetsvurdering vil arealet under kurven være 1. Figuren viser, at vi korrekt klassificerer kommunens kvalitetsvurdering på baggrund af sprogverdierne i 65 pct. af tilfældene (arealet under kurven er 0,645).

Metodefigur 1 // Klassifikation af kvalitetskategori på baggrund af sprogverdier i Københavns Kommune, 1. datasæt

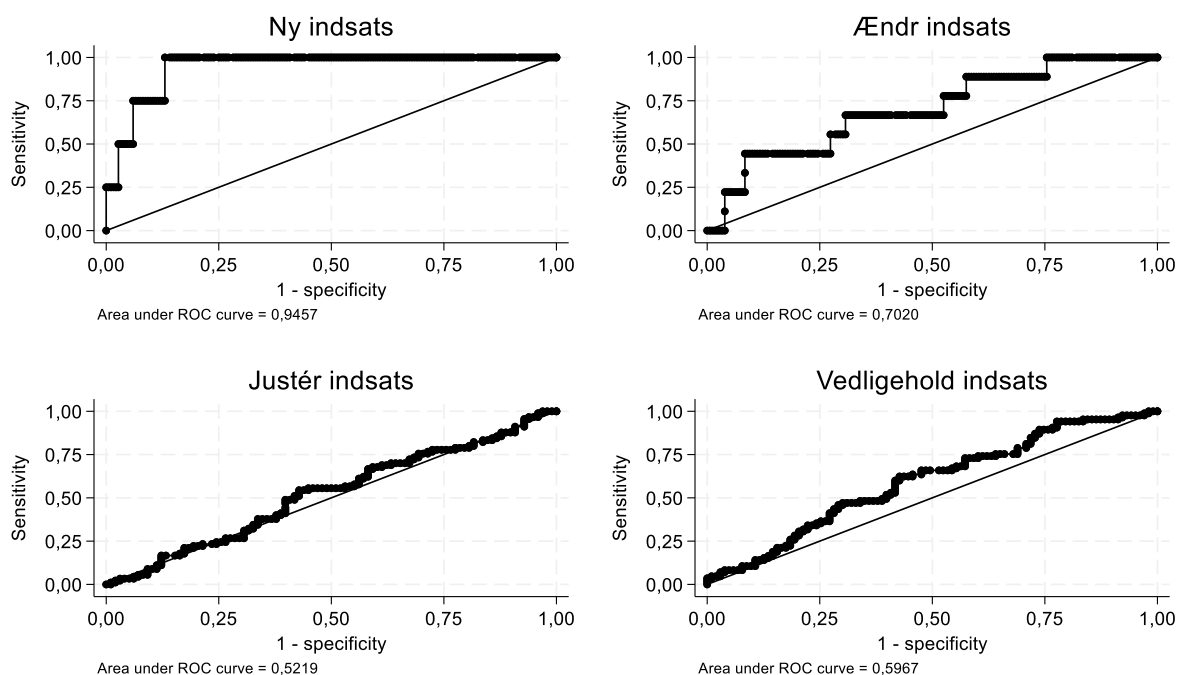


Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioden 2018 til 30.4.2022
Anm.: $n = 188$. Stikprøven er repræsentativ for den faktiske fordeling blandt alle rapporter i Københavns Kommune på et 95-pct.-signifikansniveau. Grafen plotter de sandt positive og falsk positive klassificeringer for den givne kvalitetskategori mod de resterende kvalitetskategorier på baggrund af fire uafhængige logistiske regressionsmodeller for de fire udfald (One-vs-Rest multiclass Receiver Operating Characteristic (OvR ROC)). Diagonallinjen illustrerer tilfældet, hvor klassifikationen er fuldstændig tilfældig. Arealet under kurven korresponderer med andelen af korrekt klassificerede kvalitetskategorier.

Tilsvarende kan vi opstille en model for det første pejlemærke i 2. datasæt (se metodebilag 3). Her er alle fire kvalitetskategorier repræsenteret. I hvor høj grad vi korrekt kan klassificere kvalitetskategorien, er illustreret i metodefigur 2. Vi kan korrekt klassificere den laveste kvalitetskategori på baggrund af sprogverdierne i 95 pct. af tilfældene. Rapporter med kvalitetsvurderingen "justér indsats" har vi sværest ved at skelne fra de øvrige. Her kan vi kun klassificere kvalitetskategorien korrekt baseret på sprogværdien i 52 pct. af tilfældene. Det skyldes, at der er en langt større spredning i sprogverdierne blandt de mere positive rapporter i kommunen, jf. figur 8. De to resterende kvalitetsvurderinger klassificerer vi med hhv. 70 og 60 pct. sikkerhed. Den faktiske sammenhæng mellem sprogværdier og kvalitetsvurderinger er med al sandsynlighed en del stærkere end tallene ovenfor indikerer, da sprogværdien er udregnet på baggrund af hele rapporten, mens kvalitetsvurderingen kun afspejler en sjettedel af pejlemærkerne i rapporten.

Ovenstående indikerer, at sprogværdierne afspejler kvalitetsvurderingerne for Københavns Kommune, særligt for rapporterne med de bedste og dårligste kvalitetsvurderinger. På baggrund af det, samt den stærke korrelation mellem KIDS-score og sprogværdi i metodebilag 1, finder vi det derfor rimeligt, at sprogværdierne kan anvendes som udtryk for en kommunes kvalitetsvurdering af dens daginstitutioner under de antagelser, der er udlagt i afsnittet "Begrænsninger i brugen af data".

Metodefigur 2 // Klassifikation af kvalitetskategori på baggrund af sprogværdier i Københavns Kommune, 2. datasæt



Kilde: egne beregninger på baggrund af indsamlede tilsynsrapporter i perioden 2018 til 30.4.2022

Anm.: $n = 188$. Stikprøven er repræsentativ for den faktiske fordeling blandt alle rapporter i Københavns Kommune på et 95-pct.-signifikansniveau. Grafen plotter de sandt positive og falsk positive klassificeringer for den givne kvalitetskategori mod de resterende kvalitetskategorier på baggrund af fire uafhængige logistiske regressionsmodeller for de fire udfald (One-vs-Rest multiclass Receiver Operating Characteristic (OvR ROC)). Diagonallinjen illustrerer tilfældet, hvor klassifikationen er fuldstændig tilfældig. Arealet under kurven korresponderer med andelen af korrekt klassificerede kvalitetskategorier.

Metodebilag 1. Regressionstabel: Validitet for sprogscore i forhold til KIDS-score i Slagelse Kommune, 1. datasæt.

1. datasæt, lineær regressionsmodel		
Vægtet KIDS-score	6,167 (1,841)	***
Konstantled	57,763 (4,693)	***
R2	0,296	
Antal observationer	14	

Robuste standardfejl i parentes. *** p ≤ 0,01, ** p < 0,05, * p < 0,1. I modellen anvendes en vægtet KIDS-score, se afsnittet "Regnemetoder" i metodebeskrivelsen for detaljer.

Metodebilag 2. Regressionstabel: Validitet for sprogscore i forhold til det første pejlemærke i Københavns Kommune, 1. datasæt.

1. datasæt, logistisk regression		
Sprogværdi for det første pejlemærke (Sociale relationer)		
Tilpas indsats	(Reference)	
Vedligehold indsats	0,366 (0,143)	***
Konstantled	-0,845 (0,586)	
Pseudo-R2	0,043	
Antal observationer	141	

Robuste standardfejl i parentes. *** p ≤ 0,01, ** p < 0,05, * p < 0,1.

Metodebilag 3. Regressionstabel: Validitet for sprogscore i forhold til det første pejlemærke i Københavns Kommune, 2. datasæt.

2. datasæt, multipel logistisk regression		
Sprogværdi for hele rapporten		
Ny indsats	(Reference)	
Ændr indsats	5,174 (2,067)	**
Konstantled	-10,268 (4,261)	**
Justér indsats	7,043 (1,997)	***
Konstantled	-12,677 (4,060)	***
Vedligehold indsats	7,515 (1,996)	***
Konstantled	-14,027 (4,058)	***
Pseudo-R2	0,065	
Antal observationer	188	

Robuste standardfejl i parentes. *** p < 0,01, ** p < 0,05, * p < 0,1.

Bilag 2. Regressionstabel: Persistens i sprogverdier for alle kommuner

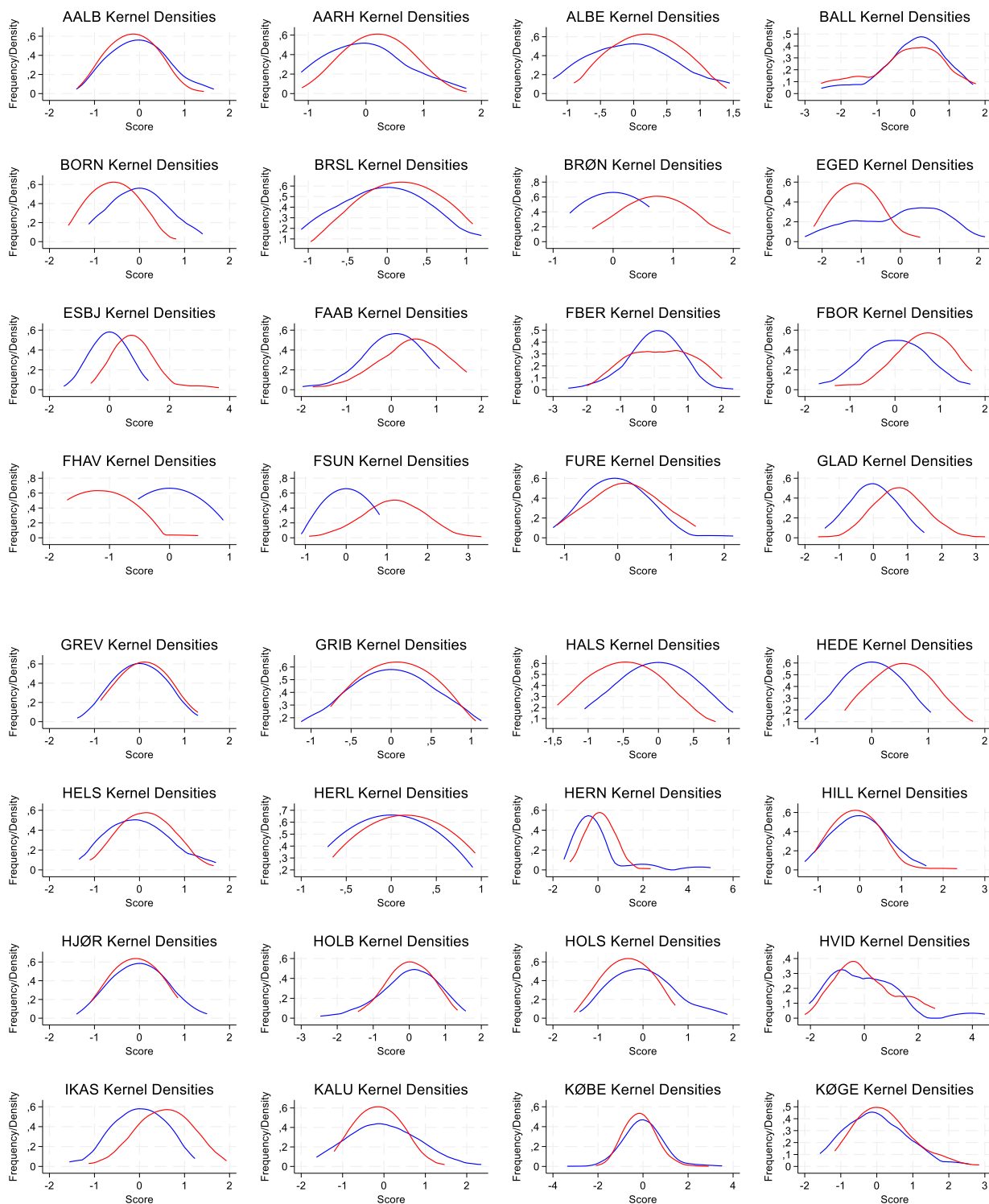
	Sprogverdi, 2. datasæt	Sprogverdi, 2. datasæt	Sprogverdi, 2. datasæt
Sprogverdi 1. datasæt	0,226 ***	0,143 ***	.
	(0,030)	(0,027)	.
Kommune (FE)	Absorberet	Absorberet	Absorberet
Margins (Kommune) i forhold til sprogverdi, 1. datasæt:			
Aalborg			0,042
			(0,059)
Aarhus			0,038
			(0,086)
Bornholm			-0,146
			(0,120)
Brøndby			-0,202
			(0,521)
Brønderslev			0,097
			(0,253)
Egedal			0,031
			(0,085)
Esbjerg			-0,101
			(0,218)
Faaborg			0,009
			(0,362)
Frederiksberg			0,260
			(0,190)
Frederikssund			0,172
			(0,566)
Furesø			0,269
			(0,265)
Gladsaxe			0,068
			(0,187)
Greve			0,005
			(0,228)
Halsnæs			-0,535 ***
			(0,150)
Hedensted			0,181
			(0,251)
Herning			0,076
			(0,051)
Hillerød			0,081
			(0,146)
Hjørring			0,019
			(0,104)
Holbæk			0,163
			(0,246)
Holstebro			0,006
			(0,073)
Hvidovre			0,263
			(0,192)
Kalundborg			-0,111
			(0,079)
København			0,221 ***

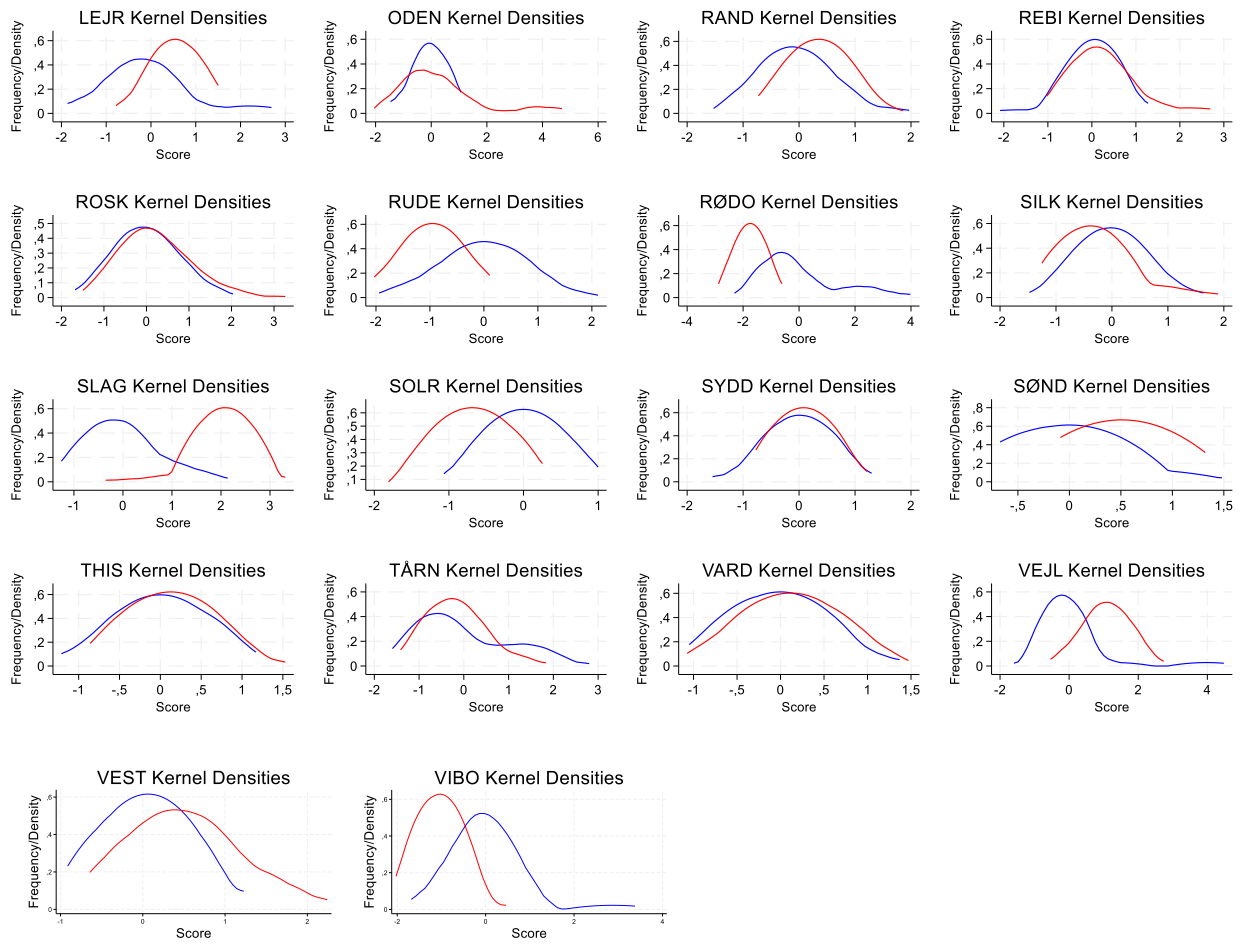
			(0,050)	
Køge			0,547 ***	
			(0,162)	
nord			0,035 ***	
			(0,006)	
Odense			2,139 ***	
			(0,683)	
Randers			0,166 **	
			(0,076)	
Roskilde			0,353 **	
			(0,145)	
Rudersdal			-0,083	
			(0,079)	
Rødovre			-0,038	
			(0,036)	
Silkeborg			0,130	
			(0,173)	
Slagelse			0,025	
			(0,080)	
Syddjurs			0,125	
			(0,178)	
Sønderborg			-0,045 **	
			(0,020)	
Tårnby			0,140	
			(0,103)	
Varde			-0,466	
			(0,277)	
Vejle			0,026	
			(0,331)	
vest			-0,131	
			(0,343)	
Viborg			0,027	
			(0,029)	
Vordingborg			0,126	
			(0,443)	
Konstantled	1,879 ***	1,444 ***	1,648 ***	
	(0,068)	(0,066)	(0,130)	
Antal observationer	1.243	1.243	1.243	

Robuste standardfejl i parentes. *** p ≤ 0,01, ** p < 0,05, * p < 0,1.

Bilag 3. Spredning i sprogverdier i alle kommuner vi kan følge over tid

I de nedenstående fordelinger er 1. datasæt blå, og 2. datasæt er rødt. Fordelinger er udregnet som i figur 11, se evt. metodebilaget. Når fordelingen for 2. datasæt fx ligger til højre for fordelingen for 1. datasæt indikerer det, at kommunens sprogverdier er steget. Her er vist fordelinger for de 49 kommuner, der har ti eller flere observationer i begge datasæt. Antal observationer, der ligger til grund for fordelingerne, samt beskrivelser af kommunekoderne er vist i bilag 5.

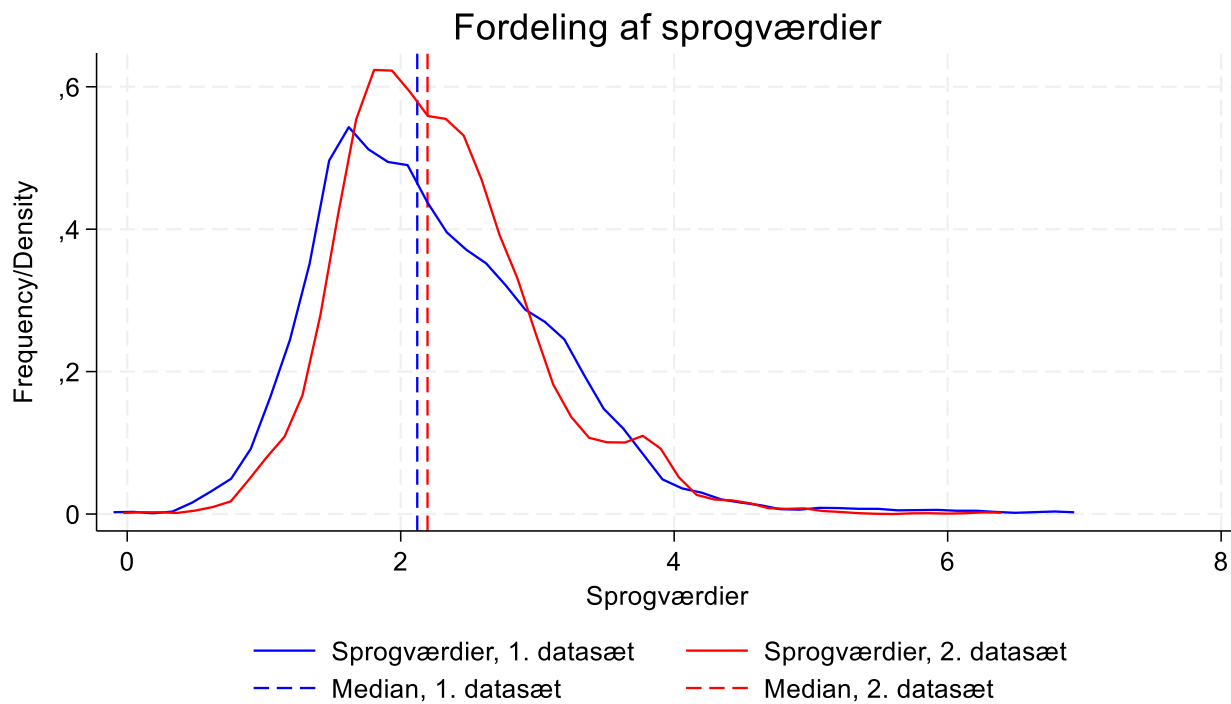




Bilag 4. Antal observationer i fordelinger i bilag 3

Kommune	Kommunekode	Antal observationer	
		1. datasæt	2. datasæt
Aalborg	aalb	106	169
Aarhus	aarh	36	178
Albertslund	albe	11	15
Ballerup	ball	27	17
Bornholm	born	18	24
Brønderslev	brsl	14	20
Brøndby	brøn	21	22
Egedal	eged	20	26
Esbjerg	esbj	67	48
Faaborg	faab	25	18
Frederiksberg	fber	81	53
Frederiksborg	fbor	13	13
Frederikshavn	fhav	18	18
Frederikssund	fsun	39	36
Furesø	fure	30	28
Gldsaxe	glad	28	51
Greve	grev	25	27
Gribskov	grib	12	18
Halsnæs	hals	11	15
Hedensted	hede	17	25
Helsingør	hels	14	14
Herlev	herl	16	22
Herning	hern	23	45
Hillerød	hill	34	41
Hjørring	hjør	42	28
Holbæk	holb	37	25
Holstebro	hols	39	39
Hvidovre	hvid	20	23
Ikast	ikas	17	18
Kalundborg	kalu	27	32
København	købe	276	368
Køge	køge	31	43
Lejre	lejr	11	16
Odense	oden	13	28
Randers	rand	31	63
Rebild	rebi	23	15
Roskilde	rosk	65	68
Rudersdal	rude	42	47
Rødovre	rødo	20	20
Silkeborg	silk	53	51
Slagelse	slag	31	41
Solrød	solr	15	14
Syddjurs	sydd	25	23
Sønderborg	sønd	46	39
Thisted	this	16	30
Tårnby	tårn	31	34
Varde	vard	24	27
Vejle	vejl	24	58
Vesthimmerland	vest	10	20
Viborg	vibo	30	47

Bilag 5: Fordeling af sprogverdier i 1. og 2. datasæt



Anm.: $n = 2.121$ og $n = 2.873$ observationer i hhv. 1. og 2. datasæt. Rapporter med en gennemsnitlig sprogverdi på over 7 er fjernet her, og det samme gælder rapporter med en længde på over 500 sætninger.



Tænk tanken DEA
Fiolstræde 44
1171 København K
www.dea.nu