

Baggrundsrapport til "Børns læsning 2017 – en kvantitativ undersøgelse af børns læse- og medievaner i fritiden"

Af Morten Rasmus Puck

Center for Anvendt Skoleforskning,

University College Lillebælt

mrpu@ucl.dk

Denne rapport vil gennemgå stikprøveudtrækningen til spørgeskemaundersøgelsen "Børns læsning 2017 – en kvantitativ undersøgelse af børns læse- og medievaner i fritiden", samt metodiske overvejelser i undersøgelsen. Spørgeskemaundersøgelsen foretages på vegne af Tænk tanken Fremtidens Biblioteker. Meningen med denne rapport er at gøre rede for de overvejelser, der ligger til grund for stikprøveudtrækningen til spørgeskemaundersøgelsen, samt hvilke analyseredskaber der er anvendt i forhold til datasættet. Hovedformålet med undersøgelsen er at undersøge børns brug af biblioteker samt læse- og medievaner, således at bibliotekerne kan tilpasse deres tilbud.

Indhold

Historikken ved undersøgelsen	4
Udfordringer ved undersøgelsen	4
Stikprøvedesign	5
Strataopdeling	5
Skolestørrelse	6
Skoletype	6
By-land skoler	6
Region	7
Stratavariablens konstruktion	8
Opgørelse af stikprøve og stratums størrelse	8
Fordelingen i stikprøverne	10
Struktur a la ICILS	12
Indsamlingsprocessen	13
Statistik over deltagende skoler	16
Deltagelsesprocenterne fra klasserne	19
Repræsentativ stikprøve	20
Vægte	24
Skoleudtrækningsvægten (SUV)	25
Skoletilsagnsvægten (STV)	25

Skoledeltagelsesvægten (SDV)	26
Elevudtrækningsvægten (EUV).....	26
Elevdeltagelsesvægten (EDV)	27
Klassedeltagelsesvægten (KLV)	27
Korrektionsvægten (Korrektion)	28
Afslutningsvægten.....	28
Totalvægte (TWGT)	29
Replikationsvægte	29
Vægte vedrørende titler og forfattere	31
Spørgeskemaets inspirationskilder	31
Rensning af data.....	32
Pilottestning af spørgeskemaet.....	33
Svarrækkefølgen.....	33
Ensretning af forfattere og titler	34
Faktoranalyser	35
Indekskonstruktion.....	45
Bibliografi	46

Historikken ved undersøgelsen

Den kvantitative undersøgelse angående børns læse- og medievaner er bestilt af Tænk tanken Fremtidens Biblioteker, og er foretaget af Læremiddel.dk sammen med Center for Anvendt Skoleforskning på University College Lillebælt. Målet er at undersøge børn i 9-13 års alderen brug af biblioteker samt læse- og medievaner, således at fremtidens biblioteker kan tilpasse deres tilbud til børnenes aktuelle brug. For at sikre et så repræsentativt udsnit af børn som muligt, har vi valgt at foretage undersøgelsen i det danske grundskolesystem. Ved at foretage undersøgelsen i skolesystemet vil vi i højere grad få spurgt eleverne som ikke normalt bruger biblioteket. Dermed vurderer vi, at den kvantitative undersøgelse vil give et godt grundlag for at kunne udtale os om børns læse- og medievaner i deres fritid. For at ramme aldersgruppen 9 til 13 år har vi valgt at sige, at vi vil spørge eleverne i 3.-7. klasse i folkeskoler og privat- og friskoler.

Undersøgelsen indeholder ikke en læsekompetencetest, og vi kan dermed ikke direkte konkludere noget om, hvorvidt elevernes læsekompetence påvirkes af deres fritidslæsning, eller omvendt. Vi vil dog senere kunne lave en teoretisk sammenkobling imellem denne undersøgelse og PIRLS 2016 undersøgelsen, således at der muligvis kan laves nogle kausale slutninger imellem fritidslæsning og læsekompetence. Dog er dette ikke inden for projektbeskrivelsen fra Tænk tanken Fremtidens Biblioteker, og er heller ikke muligt inden for tidsrammen. Dermed vil der i givet fald blive tale om sekundære analyser foretaget af Læremiddel.dk og Center for Anvendt Skoleforskning.

Udfordringer ved undersøgelsen

En af de generelle udfordringer ved kvantitative analyser i det danske grundskolesystem er, at skolerne bliver bebyrdet med undersøgelser. Dermed vil det kunne være svært at få visse af typer af skoler til at deltage. Siden vi er interesseret i hvordan billedet er på nationalt plan, har vi valgt at udtrække et fåtal af landets skoler, da det vil give os et tilstrækkeligt billede af, hvordan det forholder sig med børns læse- og medievaner i fritiden. Vi har derfor kun kontaktet de skoler, som vi mener, skal være med i undersøgelsen for at opnå repræsentativitet. En følgeudfordring heraf er, at skolerne og klasserne har haft mulighed for at melde fra til undersøgelsen, og dermed vil vi kunne have den situation, at skolerne og klasserne ikke er ligeligt repræsenteret i henhold til landsplan. Dermed må vi inddrage vægtning, således at vores undersøgelse bliver repræsentativ. Vægtningen tager højde for, at alle ikke har lige stor sandsynlighed for at deltage.

Stikprøvedesign

Desværre findes der ikke et centralt register hvori vi kan finde kontaktoplysningerne på lærerne i alle klasserne i klassetrinene 3.-7. klasse. Dermed har vi måttet få fat på lærernes kontaktoplysninger på en anden måde. Måden hvorpå at vi kan fremskaffe oplysningerne på lærerne, og dermed adgang til klasserne, er ved at tage kontakt til skolerne. For at mindske antallet af skoler vi forstyrrer, og mindske byrden af kontakt, har vi valgt at sige, at vi laver en klyngeudvælgelse, således at vi er interesseret i alle eleverne på skolen i klassetrinene 3.-7.. Det betyder at vi er interesseret i at få alle kontaktoplysningerne på alle dansklærerne i de relevante klasser på hver udvalgt skole.

En udfordring ved udvælgelsen af skolerne i klyngeformat er, at man risikerer ikke at få udtrukket skoler med bestemte karakteristika. Derfor har vi valgt lave en stratificering af skolerne ud fra fire forskellige strata/kriterier. Det vil sige, at vi i videst muligt omfang vil sikre, at vi får skoler af enhver slags med i undersøgelsen. For at sikre at omfanget rummer flest mulige slags skoler, har vi valgt at udtrække 125 skoler fordelt på folkeskoler og privat- og friskoler. Antallet af elever i efterskolen på klassetrinene 3.-7. er så lavt at vi har valgt at se bort fra dem. Ligeledes har vi set bort fra skoler med dagstilbud og specialtilbud.

Strataopdeling

Vi har valgt en udvælgelsesmetode, som sikrer et mangfoldigt udsnit af skoler i det danske grundskolesystem. De fire karakteristika, som vores udtrækningsstrata er lavet på baggrund af, er: 1) Skolestørrelse, 2) Skoletype, 3) By-land skoler, og 4) Region. Disse fire karakteristika er valgt, fordi der ligger registerdata tilgængeligt for alle skoler omkring disse forhold. Dernæst giver de en vis spredning på skolerne, og øger derved sandsynligheden for at få en repræsentativ stikprøve. Til selve opdelingen af skolerne i de forskellige kriterier har vi valgt at bruge Center for Anvendt Skoleforskningsskoledatabase, som indeholder informationer om alle skoler i Danmark. Databasen bygger på tilgængelige tal på hver skole fra Undervisningsministeriets database. Center for Anvendt Skoleforskningsskoledatabase indeholder mange oplysninger, således at vi kan lave et specifikt udtræk. For yderligere uddybninger af hvad skoledatabasen indeholder, kontakt Morten Rasmus Puck på mrpu@ucl.dk, hvorfra man vil modtage et dokument om databasens opbygning og de metodiske valg bag. Fra Center for Anvendt Skoleforskningsskoledatabase ved vi, at der findes 1564 unikke folke- eller privat- eller friskoler med enten en 3., 4., 5., 6., eller 7. klasse på skolen. Selve kriterieinddelingen vil blive angivet senere i dette dokument.

Skolestørrelse

For at danne et stratum for skolens størrelse har vi valgt at dele alle skoler op i, om det er en stor eller en lille skole. Vi har kun valgt disse to muligheder, da vi gerne vil have nogle store strata, når vi kombinerer skolestørrelsen med de andre kriterier. Det skyldes, at jo flere grupper vi får lavet, desto mindre bliver undergrupperne, og dermed må flere strata undlades, da strataet ellers kommer til at indeholde for få skoler. Vedrørende skolestørrelsen eksisterer der ikke nogle klare definitioner på, hvornår en skole er en stor skole eller ej. Vi har valgt at sige, at en stor skole har mindst 400 elever. Med mindst 400 elever på skolen er der stor sandsynlighed for at skolen har mindst ét klassetrin med 2 klasser. Det vil sige at der er stor sandsynlighed for flere hold af lærere i fag på en årgang. For at sikre at vi får det fulde billede fra hver skole, bliver alle lærerne i 3.-7. klasserne på skolen inviteret til undersøgelsen.

Når vi deler skolerne op ved skillelinjen på 400 elever på skolen, får vi, at der findes 840 små skoler med mindst en af klasserne 3.-7., og der er 724 skoler med mindst en 3.-7. klasse.

Vi har valgt at kode skolerne således, at variabelen skolestørrelse har værdien 1, hvis der er en lille skole, mens variabelen skolestørrelse har værdien 2, hvis det er en stor skole.

Skoletype

For at danne et stratum for skolens type har vi valgt at dele alle skoler op i om det er folkeskole eller en privat- eller friskole. Vi har valgt at bruge Undervisningsministeriets kategorisering af skolerne, hvorved vi relativt nemt kan få fat i oplysningerne angående skolen. Vi har valgt at inddrage skoletypen som et stratum, da der vil kunne være en systematisk forskel imellem elever i folkeskolen og på privat- eller friskolerne, som vi gerne vil prøve på at tage højde for.

Når vi deler skolerne op efter skoletype ser vi, at der er 1057 folkeskoler, som har mindst en af klasserne 3. – 7. klassetrin på skolen. Vi har, at der er 507 privat- og friskoler der har mindst en 3.-7. klasse på skolen.

Vi har valgt at kode skolerne således at variabelen skoletype har værdien 1, hvis der er tale om en folkeskole, mens variabelen skoletype har værdien 2, hvis det er en privat- eller en friskole.

By-land skoler

For at sikre en geografisk spredning imellem by- og landskoler har vi valgt at lave et stratum, der tager højde for det. Vi har valgt at opdele skolerne på nogle meget løse forhold i denne variabel. Vi har ikke

været inde på hver enkelt skoles geografiske placering for at se om skolen faktisk ligger i et byområde eller på landet. Vi har valgt at sige, at hvis skolen tilhører en kommune, som kan klassificeres som en bykommune, har skolen fået prædikatet at det er en byskole ligegyldigt hvor i kommunen skolen ligger. Vi har valgt at klassificere Aarhus, Aalborg, Odense, og hovedstadsområdet som byområder. Det vil sige, at skoler, som ligger i en af disse kommuner, vil blive betragtet som byskoler. Hovedstadsområdet består i vores undersøgelse af følgende kommuner: København, Frederiksberg, Albertslund, Brøndby, Gentofte, Herlev, Lyngby-Taarbæk, Rødovre, Tårnby, Vallensbæk, Ishøj, Greve, Ballerup, Rudersdal, Glostrup, Gladsaxe, Hvidovre og Furesø. Disse kommuner er valgt, da de ligger relativt tæt på hinanden og indre København. Grunden til at vi ser på kommuneniveau skyldes, at folkeskolerne og folkebibliotekerne tilhører kommunernes kerneområde. Kommuner, som defineres som bykommuner, vil muligvis have en anden strategi for folkebiblioteker end de kommuner, som vi definerer som landkommuner. Folkeskolerne og bibliotekerne må følge kommunens politik. Dermed må skoler liggende i landområder i bykommuner tilpasse sig de forhold der er lavet til byskolerne i bykommunerne.

Når vi deler skolerne op efter by-land skoler, er der 442 byskoler, mens der er 1122 skoler som vi definerer som landskoler.

Vi har valgt at kode by-landskole variabelen, således at skoler der bliver betragtet som landskoler har fået værdien 1, mens skoler der er betragtet som byskoler har fået værdien 2.

Region

Den sidste opdeling vi har lavet, er hvor i landet skolen er placeret. Vi har valgt at lave strataet via hvilken region skolen ligger i. Vi har valgt regionniveau, og ikke kommuneniveau, fordi vi vil få større grupper, idet flere skoler vil leve op til de fastlagte karakteristika. Hermed er der også flere muligheder for at udvælge skoler inden for grupperne. Derudover har skoler der ligger i samme region på sin vis lignende forhold, at arbejde under. Vi minder om, at vi allerede har lavet by-land variabelen, som vil dele de fleste regionstrata op i en land- og by-del. På denne måde øger vi sandsynligheden for, at vi vil få udvalgt skoler spredt ud over regionerne og dermed vil hele landet geografisk kunne blive repræsenteret.

Vi har valgt, at der ligger 346 skoler i region Hovedstaden, 225 skoler i region Sjælland, mens der er 378 skoler i region Syddanmark, 436 skoler i region Midtjylland, og 179 skoler i region Nordjylland. Fælles for alle skolerne er, at de har mindst en klasse på klassetrinene 3., 4., 5., 6, eller 7. klasse. Vi har valgt, at hvis en skole ligger i region Hovedstaden har vores regionvariabel værdien 1. Hvis skolen ligger i region Sjælland, så har skolen fået værdien 2 i vores regionvariabel. Skulle skolen ligge i region Syddanmark vil

skolen have værdien 3 i regionvariablen, mens hvis skole lå i region Midtjylland vil skolen have værdien 4 i regionvariablen. Til sidst vil en skole i region Nordjylland have værdien 5 i vores regionvariabel.

Stratavariablens konstruktion

For at kunne lave udtrækningen nemmest i statistikprogrammet SAS, har vi valgt at kombinere værdierne fra de fire strata til en unik firecifret stratakode. Rækkefølgen på den unikke kode er: Skolestørrelse, Skoletype, By-land skoler, og Region. Ud fra de unikke koder kan vi sige hvilken type skolen er, blot ved at betragte deres stratums unikke kode. Et eksempel kunne være 1224, som vil være en lille privat/friskole i byen i region Midtjylland.

Opgørelse af stikprøve og stratums størrelse

I Tabel 1 har vi opskrevet alle de mulige skolestrata i datasættet, der er relevant for vores undersøgelse. Når vi i dette afsnit taler om skoler, mener vi kun de skoler, som har mindst en klasse i enten 3., 4., 5., 6. eller 7. klassetrin. For hvert strata har vi ligeledes optalt hvor mange skoler der er i populationen i det pågældende stratum, samt hvor mange skoler fra hvert strata vi har udtrukket. Vi har i alt 40 mulige strata som skolerne kan udtrækkes fra. Vores stikprøvemål på 125 skoler er fordelt på strata, således at fordelingen i stikprøven minder om fordelingen i populationen. Da vi udtrækker skolerne i stikprøven som landets skoler er fordelt, vil nogle strata have for få skoler til at der bliver udtrukket en skole derfra. I Tabel 1 kan vi se at strata 1122 ikke har fået udtrukket nogen skoler, mens strata 1111 har fået udtrukket en enkelt skole. Det skyldes at stratum 1122 ikke har skoler nok til at leve op til at have mindst 0,4% af landets skoler i sig. De 0,4% kommer fra afrunding til nærmeste hele antal skole i stikprøven. Der er i alt 33 strata, som har mindst 0,4% af alle landets skoler i sig. Det betyder at 7 strata forsvinder, og dermed er der 18 skoler, som ikke har mulighed for at deltage i undersøgelsen, da der ikke blev udtrukket nogen skoler fra deres stratum. De 18 skoler udgør samlet 1,15% af alle landets skoler.

Det vigtigste formål med en stikprøve er at få reduceret usikkerheden på estimerne videst muligt, samt at minimere omkostningen ved dataindsamlingen. Dette gøres ved at udtrække de grupper, der indeholder størst varians og få kortlagt variansen, således at der bliver mindre usikkerhed på estimerne. Variansen indenfor grupperne er der desværre ikke noget tilgængelig empiri på, således har vi valgt at udtrække grupperne efter bedste bud, hvilket vil være en fordeling imellem strataerne, som populationen er. Fremadrettet vil man kunne bruge denne undersøgelse til at finde ud af hvordan man kan lave en bedre stikprøve, når man skal undersøge elevernes læse- og medievaner i fritiden.

Tabel 1 viser os fordelingen af skoler inden for de forskellige strata. Både i form af populationen og stikprøven. Vi har også tilføjet hvert enkelt stratums andel af skoler i hele populationen.

Tabel 1: Fordelingen af skoler i populationen og stikprøven

strata	Populations antal	Populations andel	Stikprøves antal
1111	7	0,4%	1
1112	47	3,0%	4
1113	101	6,5%	8
1114	158	10,1%	12
1115	38	2,4%	3
1121	12	0,8%	1
1122	2	0,1%	0
1123	7	0,4%	1
1124	8	0,5%	1
1125	27	1,7%	2
1211	27	1,7%	2
1212	72	4,6%	6
1213	106	6,8%	8
1214	85	5,4%	7
1215	39	2,5%	3
1221	69	4,4%	5
1222	1	0,1%	0
1223	13	0,8%	1
1224	11	0,7%	1
1225	10	0,6%	1
2111	49	3,1%	4
2112	85	5,4%	7
2113	113	7,2%	9
2114	120	7,7%	9
2115	38	2,4%	3
2121	150	9,6%	12
2122	8	0,5%	1
2123	26	1,7%	2

2124	38	2,4%	3
2125	23	1,5%	2
2211	8	0,5%	1
2212	9	0,6%	1
2213	8	0,5%	1
2214	10	0,6%	1
2215	2	0,1%	0
2221	24	1,5%	2
2222	1	0,1%	0
2223	4	0,3%	0
2224	6	0,4%	0
2225	2	0,1%	0

Note 1: Tallene kommer fra udtrækning fra CAS' skoledatabase

Fra Tabel 1 kan man se de 7 strata, hvor der ikke er udtrukket nogen skoler fra. Som sagt dækker de 7 strata kun over 18 skoler, som der ikke udtrækkes fra. Primært drejer det sig primært om store privat- og friskoler i byerne. Umiddelbart vurderer vi ikke at det vil skævvride vores analyser, men vi har valgt at lave en kontrol af stikprøvefordelingen med populationsfordelingen.

Fordelingen i stikprøverne

Siden vi har med en stikprøve at gøre, kan vi ikke vide os sikre på, at vi har en fornuftig stikprøve, før vi har sammenlignet vores stikprøves karakteristika med populationens karakteristika. Vi har derfor valgt at se på, hvordan andelen i forskellig henseende forholder sig i stikprøven og populationen. Vi har valgt at se på vores udtrukne skoler, og ikke de deltagende skoler, da vi primært er interesseret i at se om vores stikprøveudtrækning forfordeler en bestemt gruppe. Da vi har udtrukket skolerne fra specifikke skolestrata, vil der summarisk kunne ske en skævvridning, hvis der tilfældigvis bliver udtrukket for mange skoler af en speciel type. Kriterierne for opdelingen af de enkelte variable kan læses tidligere i denne rapport.

Vi starter med at vurdere om skolestørrelsesvariablen i vores stikprøve passer med andelens størrelse i populationen.

Tabel 2: Sammenligning af udtrukne skoler i stikprøven og populationen i henhold til skolestørrelse

	Antal	Andel stikprøve	Andel population
Lille	67	54%	54%
Stor	58	46%	46%

Note 2: Egne beregninger

Som det kan ses i Tabel 2 passer fordelingen i stikprøveudtrækket med fordelingen i populationen. 54% af alle skolerne i stikprøven har færre end 400 elever på skolen. Det svarer til den samme andel i populationen.

Vi ser nu på fordelingen af skoletype i vores stikprøve

Tabel 3: Sammenligning af udtrukne skoler i stikprøven med populationen i henhold til skoletype

	Antal	Andel stikprøve	Andel population
Folkeskole	85	68%	68%
Privat- og friskoler	40	32%	32%

Note 3: Egne beregninger

Som det kan ses i Tabel 3 er fordelingen af skoletyperne i stikprøven helt ens med fordelingen i populationen. Dermed vil stikprøven kunne give et retvisende billede af hvordan læselysten er for børn i 3.-7. klasse i det danske grundskolesystem, da der ikke er nogen udtrækningskævringer via skoletype, hvilket kunne tænkes at have den største effekt på resultaterne.

Vi ser nu på fordelingen af by-land skoler i vores stikprøve.

Tabel 4: Sammenligning af udtrukket skoler i stikprøven med populationen i henhold til by-land skoler

	Antal	Andel stikprøve	Andel population
Landskoler	90	72%	72%
Byskoler	35	28%	28%

Note 4: Egne beregninger

Som det kan ses i Tabel 4 er fordelingen af by-land skoler ens med fordelingen i populationen.

Vi ser nu på fordelingen af den regionale placering af skolen.

Tabel 5: Sammenligning af udtrukket skoler med populationen i henhold til regional placering

	Antal	Andel stikprøve	Andel population
Region Hovedstaden	28	22%	22%
Region Sjælland	19	15%	14%
Region Syddanmark	30	24%	24%
Region Midtjylland	34	27%	28%
Region Nordjylland	14	11%	11%

Note 5: Egne beregninger

Som det kan ses i Tabel 5 med fordelingen af skolernes regionale placering, er der overvejende overensstemmelse mellem stikprøvens fordeling og populationens fordeling. Forskellene er så små, at de udelukkende skyldes afrundinger.

Alt i alt vil vi betragte vores fordeling af udtrukne skoler i stikprøven som et repræsentativt udsnit af alle skoler i det danske grundskolesystem med klasser i 3.-7. klasse. Dermed er vi sikre på at alle typer af elever og børn vil kunne blive repræsenteret i undersøgelsen. Det betyder at vi vil ende med et generaliserbart resultat, hvis skoler fra alle udtrukne strata deltager.

Årsagen til at vi ser på fordelingen af udtrukne skoler og ikke de deltagende skole er, at vi vil vægte de deltagende skoler, således at de kommer til at ligne fordelingen af de udtrukne skoler. Det betyder at vores datasæt vil opføre sig som de udtrukne skoler, og derfor er det at foretrække at de udtrukne skoler er repræsentative.

Struktur a la ICILS

Strukturen i vores udtrækningsproces for skolerne har fulgt lignende internationale undersøgelser, såsom ICILS, TIMSS, PIRLS og PISA. Det betyder, at vi ikke bare har udtrukket 125 skoler og kontaktet dem. Vi udtrækker først 125 skoler, som vi kalder for samlede skoler (udtrukne skoler), og derefter udtrækker vi erstatninger for disse samlede skoler. Vi har udtrukket to erstatningsskoler for hver udtrukket skole, hvor de først udtrukne erstatningsskoler kaldes for *first replacement skoler*, og de sidst udtrukne skoler kaldes for *second replacement skoler*. Det er vigtigt, at skolerne minder om hinanden, og derfor vi har valgt at udtrække efter samme fordelingsnøgle i de tre udtrækningsrunder. Dernæst har vi sat de tre runder sammen, således at hver række i vores oversigt består af tre "ens" skoler, dvs. de tre skoler kommer fra

samme unikke strata. Hver række vil bestå af den samlede skole, en first replacement skole, samt en second replacement skole. Vi er overbevist om at skolerne minder om hinanden, da de har de samme overordnede karakteristika, fordi de kommer fra en af de 33 mulige strata, hvor vi udtrækker skoler fra.

Ved brug af stratificeret klyngeudvælgelse, som skoleudtrækningsstrategien er, er der reelt kun brug for én skole per stratum for at kunne lave en meningsfuld og repræsentativ analyse. Vi kontakter som udgangspunkt 125 skoler, men har reelt kun brug for deltagelse fra mindst én skole fra hvert af de 33 strata, der er udtrukket skoler fra. Der er valgt 125 folkeskoler og privat- og friskoler, da vi vurderer det til at være et tilstrækkeligt antal til at skabe en fyldestgørende stikprøve for elever i 3.-7. klasse i det danske grundskolesystem. Vi har valgt antallet 125, da andre lignende undersøgelser har brugt henholdsvis 100 folkeskoler i en rapport om nationale test (Bundsgaard & Puck, 2016), og 150 folkeskoler, privat- og friskoler, og efterskoler i en større læremiddelundersøgelse (Rusmann, Ahlmann Plesner, & Puck, 2017). Siden Børns læsning 2017-undersøgelsen vedrører folke- og privat- og friskoler, vil det være nærliggende at følge antallet af skoler fra læremiddelundersøgelsen (undtagen efterskolerne). Dermed rammer man et skoleantal på 125 fordelt blandt folkeskoler og privat- og friskoler fordelt over hele landet. Med et skoleantal på 125 sikrer vi også at strata med få antal skoler, vil have en sandsynlighed for at blive udtrukket. Det gør sig blandt andet gældende for små folkeskoler på landet i region Hovedstaden.

Den valgte struktur med en samlet, first replacement, og second replacement skoler i hver række, har yderligere den fordel at man ved hvor man skal fokusere kræfterne i indsamlingen af skoler. Det skyldes, at når en skole gerne vil deltage i undersøgelsen, kontakter man ikke de resterende skoler i rækken, og man dermed tager vi kontakt til færre skoler, end det ville være tilfældet hvis vi havde kontaktet alle udtrukne skoler. Det er ligegyldigt for undersøgelsens validitet om de deltagende skoler er samlede, first replacement eller second replacement. Hvis den samlede skole ikke ville deltage, gik vi videre til den næste i rækken, dvs. first replacement skolen. Derudover vil strata med hidtil ingen deltagere blive fremhævet, således man kan fokusere flere kræfter på at få en af de udtrukne skoler fra strataet med i undersøgelsen, da det er vigtigt at få mindst en skole med fra hvert stratum. Vi har således øget sandsynligheden for, at vi får en deltagende skole i hvert stratum. Generelt er det lykkedes med at indsamle skoler, men enkelte gange måtte vi udtrække en third replacement skole i strata samt enkelte fourth replacement skoler. Selve indsamlingsprocessen vil vi uddybe i næste afsnit.

Indsamlingsprocessen

Selve indsamlingsprocessen af svarene fra eleverne har været igennem flere processer inden vi kunne få selve svarene fra eleverne. Vi ville gerne have styr på hvilken klasse de enkelte elever gik i, og således har vi først skulle have fat på klassens dansklærer, for at vi kunne få spørgeskemaet afviklet i en af dansklærernes

timer. Desværre findes der ikke et centralt register med lærernes kontaktoplysninger, og vi måtte derfor gennemgå en proces, hvor vi kontaktede skolerne for at få kontaktoplysningerne på lærerne. Udvælgelsen af skolerne har fulgt den stratificerede klyngeudvælgelse, og vi har dermed sikret os, at vi har skoler af forskellige typer fra hele landet.

Kontakten til skolerne bestod i første omgang af et udskrevet brev til de samlede skolers læsevejledere udsendt i uge 41 2016. I brevet skrev vi hvad undersøgelsen gik ud på, og hvad vi skulle bruge fra administrationens side af. Siden Børns læsning 2017-undersøgelsen omhandler børn i alderen 9-13 år, hvilket svarer til 3.-7. klasse, har vi derfor bedt om kontaktoplysningerne på alle dansklærerne i 3.-7. klasse. Fra administrationens side har vi bedt om at få navnet på læreren, klassen læreren underviser, samt e-mail-adressen på læreren. Grunden til at vi vil have e-mail-adressen på læreren er, at vi vil udsende spørgeskemaet via selvoprettende link fra Survey-Xact (et program til at gennemføre spørgeskemaundersøgelser elektronisk).

Hvis skolen ikke meldte tilbage inden for en uge i form af et afbud til deltagelse i undersøgelsen eller i form af en liste med kontaktoplysningerne på skolens lærere, fik skolen en rykkermail. Rykkermailen blev sendt til skolens officielle e-mail-adresse. I denne undersøgelse fik alle samlede skoler rykkermails indtil skolen meldte tilbage om de vil deltage eller ej. Først da blev de fjernet fra kontaktoplysningslisten. Hvis skolen tilmeldte sig undersøgelsen, afsluttede vi indsamlingen i den pågældende række, hvilket vil sige at vi ikke tog kontakt til skolerne, der stod som de næste i rækken. Hvis den samlede skole meldte fra, tog vi kontakt til first replacement skolen. Denne proces fortsatte indtil vi var nødt til at igangsætte spørgeskemainsamlingerne. Vi ved fra erfaring fra andre undersøgelser, at skolerne har mest tid til at deltage i spørgeskemaundersøgelser i januar og februar. Vi besluttede derfor at undersøgelsen skulle afholdes i perioden 15/1-28/2 2017. Fordi vi havde besluttet, at undersøgelsen skulle starte i medio januar, måtte vi tage kontakt til næste led i rækken af skoler, selvom den samlede skole ikke havde afgivet noget svar trods flere rykkere. Det betød, at vi risikerede at kunne få tilsagn fra to skoler i en række, ved at first replacement skolen tilmeldte sig før den samlede skole. Det er i alt sket 12 gange. Vi har valgt at inddrage alle tilmeldte skoler. Det har dog begrænset betydning, da vi alligevel korrigerer for det i vægtningen af de enkelte observationer. Selve vægtningsprocessen vil blive forklaret senere hen i rapporten.

Den første kontakt blev som nævnt skabt via et brev sendt i uge 41. Derefter blev brevet fulgt op af en rykkermail i uge 44 og en i uge 45. Dette fortsatte indtil medio november 2016, hvorefter first replacement skolerne for de samlede skoler, der ikke havde svaret, blev kontaktet. Derfra blev både de adspurgte samlede og first replacement skoler rykket indtil der kom et svar fra en af skolerne. I medio december blev second replacement skolerne kontaktet for de rækker, hvor hverken de samlede eller first

replacement skolerne havde tilmeldt sig. Vi tog telefonisk kontakt til 7 skoler i primo januar, da vi manglede skoler fra de strata som de 7 skoler tilhører.

Et vigtigt element, når man har foretaget en stratificering i udvælgelsen af ens stikprøve, er, at man sikrer sig, at man får mindst en deltager fra hvert stratum. I denne undersøgelse har vi brugt meget specifikke strata, hvilket betyder, at der vil være få observationer i hvert stratum. Generelt har vi fået et tilsagn fra mindst en skole i hvert stratum fra enten den samlede skole, first replacement eller second replacement skolerne. Vi har problemer med at indsamle skoler som er små folkeskoler på landet i region Hovedstaden, små privat- eller friskoler på landet i region Hovedstaden, små folkeskoler i byen i region Hovedstaden, samt små privat- eller friskoler i byen i region Midtjylland. For disse strata har vi udtrukket third og fourth replacement skoler, således at vi fik en skole fra det pågældende stratum. Når vi har fået mindst én skole med fra hvert stratum, betyder det at analyserne vil blive mere retvisende ud fra et repræsentativitets perspektiv. Ingen skoler fra et stratum vil betyde, at vi ikke vil få en repræsentativ analyse.

Da vi havde indsamlet tilstrækkelig med kontaktoplysninger på lærerne på de udtrukne skoler, påbegyndte vi forberedelserne til selve undersøgelsen blandt eleverne. Vi har valgt, at hver klasse skulle have sit unikke eksemplar af spørgeskemaet, da det vil gøre det væsentlig nemmere at konstruere elev- og klassevægtene senere hen. Efter vi havde fået kontaktoplysningerne på lærerne fra skolen, kontaktede vi lærerne og bad dem om at bekræfte, at de var lærer i den pågældende klasse og på den pågældende skole. Vi spurgte også lærerne hvor mange elever der var i hver klasse, og hvornår de forventede at kunne tage testen. Den oplyste testdato var kun vejledende, da linket til spørgeskemaundersøgelsen var åben fra lærerne modtog linket og indtil at vi lukkede undersøgelsen i primo marts. Det oplyste elevtal skulle vi bruge til at skabe elevvægtene; hvad det indebærer vil blive forklaret længere nede. En fordel ved denne mailkorrespondance imellem os og lærerne er, at vi får bekræftet, at vi har den rette mailadresse, og at læreren er interesseret i at deltage i undersøgelsen. Dette øger validiteten af undersøgelsen, da vi dermed ville have mulighed for at reducere antallet af fejl og manglende deltagelse inden undersøgelsen blev afsluttet.

Som nævnt tidligere blev undersøgelsen planlagt til at forløbe i perioden fra 15/1 - 28/2 2017. Dog blev undersøgelsen forlænget, således at undersøgelsen reelt sluttede den 8/3, da enkelte lærere først kunne deltage efter den 1/3. På den måde havde vi også mulighed for at tage kontakt til alle de lærere der manglede at deltage, selvom de havde oplyst, hvornår de vil tage testen. Dette øgede antallet af deltagende klasser med cirka 20. Langt de fleste skoler har deltaget via computer, hvor klasselærerne havde fået et spørgeskemalink til klassen, som klassens elever, så kunne gå ind og besvare. Dog ville to skoler kun deltage via papir. Disse to skoler fik tilsendt en mail med en PDF-version af spørgeskemaet,

hvorefter skolerne printede spørgeskemaerne ud. Vi fik dem tilbage som skanninger via mail eller som pakkepost. Vi satte en studentermedhjælper til at skrive elevbesvarelsener ind på de respektive spørgeskemalink. Vi vurderer, at det ikke har påvirket datakvaliteten af vores undersøgelse ved at to skoler besvarede via papir. Det blev en nødvendighed for os, da en af skolerne lå i et stratum, som vi ikke havde fået deltagelse fra før. Ved at de besvarede på papir, blev undersøgelsen landsdækkende.

Statistik over deltagende skoler

Vi vil nu se på statistikken over de skoler, der har deltaget i undersøgelsen. Vi kunne have set på hvilke skoler, der har givet tilsagn om at de gerne vil deltage i undersøgelsen. Det vil sige at skolerne har sendt kontaktoplysninger på lærerne. Vi har valgt kun at se på fordelingen af deltagende skoler, da 5 skoler har givet tilsagn om at de gerne vil deltage i undersøgelsen, men ingen af skolernes klasser har deltaget i undersøgelsen. Enten ved at lærerne ikke har vendt tilbage på vores mail angående bekræftelse, elevantal, og testpunkt, eller at lærerne ikke har lavet undersøgelsen alligevel, selvom de havde alt hvad de skulle bruge for at kunne lave undersøgelsen.

Når vi ser på de deltagende skoler, ser vi på om der er en systematik i hvordan skolerne og lærerne melder tilbage og deltager. Hvis der er en systematik, kan man bruge denne oplysning i fremtidige stikprøveundersøgelser, således at man kan foretage en bedre stikprøveudtrækning. Hvis vi ved at store skoler er sværere at få fat i, betyder det at vi skal udtrække flere store skoler fremadrettet. Jo mere vi ved om en specifik type skole, desto mere præcise bliver vores estimater samlet set.

Inden vi går ind i fordelingen af de deltagende skoler, og deres karakteristika, minder vi om koderne for forskellige stratakategorier.

Tabel 6: Kodningen af de enkelte skoleopdelinger

	Skolestørrelse	Skoletype	By-land	Region
1	Lille	Folkeskole	Land	Hovedstaden
2	Stor	Privat- og friskole	By	Sjælland
3	-	-	-	Syddanmark
4	-	-	-	Midtjylland
5	-	-	-	Nordjylland

Note 6: Opdelingerne sker ved egne definitioner forklaret tidligere

I vores stikprøve over skoler har de 110 skoler, der har valgt at deltage, og de fordeler sig således.

Tabel 7: Fordelingen af deltagende skoler i de forskellige kategorier

	Skolestørrelse	Skoletype	By-land	Region
1	61	77	80	24
2	49	33	30	12
3				28
4				30
5				16

Note 7: Egne optællinger

I Tabel 7 kan man se, hvordan de deltagende skoler fordeler sig i denne undersøgelse. Tallene yderst til venstre angiver hvilken kodning de forskellige variable har fået tidligere. Man kan se de overordnede tendenser, men vi kan ikke rigtig vurdere det i forhold til vores oprindeligt udtrukne skoler. Dette skyldes at der kun deltager 110 skoler, mens der var udtrukket 125 skoler i datasættet. I de næste tabeller kan vi se hvordan fordelingen er i andele.

Tabel 8: Fordelingen af deltagende skoler i procent i henhold til de forskellige kategorier

	Skolestørrelse	Skoletype	By-land	Region
1	55%	70%	73%	22%
2	45%	30%	27%	11%
3				25%
4				27%
5				15%

Note 8: Egne udregninger

I Tabel 8 kan vi se den forholdsvise fordeling af alle kategorityper. Vi kan se at 55% af alle deltagende skoler i undersøgelsen er små skoler, hvilket vil sige skoler der har færre end 400 elever samlet på skolen. Fordelingen virker ganske fornuftig, men det er svært at vide, hvordan det forholder sig til populationens fordeling. Vi har derfor fundet, at fordelingen af de udtrukne skoler, som vi fandt, var repræsentativ for populationen, og vi kan dermed sammenligne de to fordelinger af skoler.

Tabel 9: Fordelingen af de udtrukne skoler i procent i henhold til de forskellige kategorier

	Skolestørrelse	Skoletype	By-land	Region
1	54%	68%	72%	22%
2	46%	32%	28%	15%
3	-	-	-	24%
4	-	-	-	27%
5	-	-	-	11%

Note 9: Egne udregninger

I Tabel 9 kan vi se at 54% af alle skoler i vores population er små skoler. For at vi bedre kan se om der er en forskel imellem de deltagende skoler og de udtrukne skoler har vi valgt at finde forskellen i procentpoint imellem de deltagende skoler og de udtrukne skoler i undersøgelsen.

Tabel 10: Forskellen i procentpoint i mellem de deltagende skoler og de udtrukne skoler

	Skolestørrelse	Skoletype	By/land	Region
1	2%	2%	1%	-1%
2	-2%	-2%	-1%	-4%
3	-	-	-	1%
4	-	-	-	0%
5	-	-	-	3%

Note 10: Egne udregninger. Afrundinger gør at summen af forskellene ikke nødvendigvis er 0.

Tabel 10 skal læses på den måde, at når der er et positivt tal, så er der en højere andel af de deltagende skoler, der er af denne type, end den udtrukne andel siger, at der skal være. Hvis tallet er stort, så betyder det har været relativt let at få skolerne af denne type til at deltage. I Tabel 10 kan vi se at det overordnet har været lige let at få fat i skolerne. Forskelle under 2% point er uden betydning. Vi kan dog se, at det har været relativt svært at få skoler fra Region Sjælland med, mens det har været relativt lettere at få skoler fra region Nordjylland med i undersøgelsen. Forskellene er dog relativ små, og umiddelbart vil det heller ikke få nogen betydning for vores estimater, da vi vægter de forskellige grupper op, således at grupperne fra de deltagende skoler kommer til at minde i størrelse om populationens fordeling. Det betyder, at de nordjyske skoler vil få en relativt mindre vægt end skoler fra region Sjælland. På den måde vil den vægtede fordeling minde om populationens fordeling.

Deltagelsesprocenterne fra klasserne

Via vores oplysninger fra indsamlingsprocessen kan vi udregne deltagelsesprocenten for undersøgelsen og klasserne.

Tabel 11: Fuldførelsesprocenten i undersøgelsen

	Antal fuldførte besvarelser	Antal mulige besvarelser	Fuldførelsesprocenten
Stikprøven	8721	10181	85,6%

Note 11: Egne optællinger og udregninger

I Tabel 11 kan vi se at hele undersøgelsen har en fuldførelsesprocent på 85,6%, hvilket betyder af alle dem vi har spurgt om at deltage har over 80% deltaget. Det betragtes som meget tilfredsstillende for undersøgelsen.

Vi har udregnet deltagelsesprocenten fra alle deltagende klasser for at finde ud af, om vi har en god svarprocent i klassen. Vi har som sagt spurgt hver lærer om, hvor mange elever der er i hver klasse, når alle eleverne er til stede. Ud fra det oplyste elevantal i klassen og antal af respondenter på klassens link har vi kunne udregne deltagelsesprocenter for klasserne.

Tabel 12: Tabel oversigt deltagelsesprocenter i klasserne.

Deltagelsesprocentsinterval	Andel af klasser
[25%-35%[0,4%
[35%-45%[0,7%
[45%-55%[1,1%
[55%-65%[2,4%
[65%-75%[10,6%
[75%-85%[30,2%
[85%-95%[37,0%
[95%-105%[13,4%
[105%-115%[1,8%
[115%-125%[0,2%
[125%-135%[0,0%
[135%-145%[0,4%
[145%-155%[0,4%

[155%-165%[0,0%
[165%-175%[0,7%
[175%-185%[0,4%
[185%-195%[0,2%

Note 12: Egne beregninger ud fra antal deltagende elever og det oplyste elevantal.

20 klasser har en deltagelsesprocent over 100%. Det drejer sig om 4,4% af alle deltagende klasser, dvs. et fåtal af det samlede antal klasser. Umiddelbart kan det skyldes, at nogle lærere har misforstået de tilsendte informationer, og dermed brugt et andet link end det link der blev sendt til den specifikke klasse. Det stod eksplicit i instruktionerne, at linket kun måtte bruges til besvarelse i den tildelte klasse. For disse få klassers vedkommende, kan vi desværre ikke adskille de forskellige klassers besvarelser, da vi ikke ved hvilke andre klasser der har besvaret spørgeskemaet. At der er flere besvarelser end det oplyste elevtal, gør intet, da vægtningsprocessen korrigerer de enkelte elevers besvarelser, således at eleverne, der har besvaret klassens link, kun vejer som det oplyste elevtal. Har vi fx fået oplyst at der er 22 elever i den pågældende klasse men klassens samlede antal besvarelser er 44, vil den enkelte elev kun vægte $\frac{1}{2}$ i den endelige opgørelse. Det betyder, at elever i klasser med en deltagelsesprocent mindre end 100% vejer mere end 1 person, mens klasser med en deltagelsesprocent større end 100% vejer mindre end 1.

For nogle af de 20 klasser kan svarprocenten over 100% skyldes, at vi har fået oplyst et klasseantal lavere end det reelle elevantal. Det kan skyldes, at der kommet en elev ind i klassen i perioden fra vi har fået oplyst klasseantallet til afholdelsen af testen. En yderligere mulighed er, at nogle elever har besvaret spørgeskemaet flere gange. Igen har det ingen betydning for vores estimater senere hen, da de klassers elever bliver vægtet efter deltagelsesprocenten.

Alt i alt er vi overordnet tilfredse med vores deltagelsesprocenter, og vi vurderer at deltagelsesprocenterne er tilstrækkeligt store til, at de vil give et repræsentativt billede af, hvordan det forholder sig i klassen. Vi ser i vores undersøgelse, at 91,2% af alle deltagelsesprocenterne i klasserne ligger i mellem 65%-105%. Det betyder, givet deltagelsesprocenterne og vores udtrækningsdesign, at vi får et repræsentativt billede for læsningen i fritiden blandt børn i alderen 9-13 år. For at bekræfte dette, har vi i det følgende afsnit vurderet vores stikprøve i henhold til nogle karakteristika for populationen hentet fra Danmarks Statistik.

Repræsentativ stikprøve

For at vurdere, om vi har med en repræsentativ stikprøve at gøre, er vi nødt til at vurdere stikprøven i henhold til et register. Desværre har vi ikke adgang til et direkte register, som indeholder alle de variable vi

er interesserede i. I stedet vil vi bruge oplysninger fra Danmarks Statistik til at vurdere om vores stikprøve er repræsentativ.

Siden vi har med stratificeret klyngeudvælgelse, hvor alle ikke har lige stor sandsynlighed for at deltage, vælger vi at vurdere på den vægtede stikprøve i henhold til det centrale register. På denne måde kan vi se om vores vægtede stikprøveresultater vil være repræsentative for vores population.

Vi starter med at sammenligne kønsfordelingen i stikprøven med kønsfordelingen i populationen. Vi har taget tallet fra Danmarks Statistiks tabel UDDAKT20.

Tabel 13: Kønsfordelingen i henholdsvis stikprøven og populationen

	Stikprøve	Population
Drenge	49,5%	51,1%
Piger	50,5%	48,9%

Note 13: Egne optællinger på vægtede data. Populationsdata kommer fra DST.dk's tabel UDDAKT20

Som det kan ses i Tabel 13 er kønsfordelingen i stikprøven relativt tæt på kønsfordelingen i populationen, men vi har lidt færre drenge, som deltager i undersøgelsen kontra hvor mange drenge der er i populationen. Der kan være et par forklaringer på hvorfor det kan være tilfældet. For det første er populationstallet af Danmarks Statistik opgjort den 1. oktober 2016. Siden vores undersøgelse er foretaget i januar og februar 2017, vil det være en mulighed at nogle af børnene er rykket til en anden skoletype end folkeskole eller privat- og friskole. Da undersøgelsen kigger på klassetrinene 3. til 7. klasse, vil ingen elever skifte til en efterskole, og dermed kan eleverne kun skifte til specialskoler. Hvis det primært er drenge der skifter fra folkeskoler og privat- og friskoler til specialskoler, og at de er skiftet efter den 1/10, betyder det at andelen af drenge i populationen i januar og februar 2017 vil være lavere end det opgjorte hos Danmarks Statistik. Det betyder, at undersøgelsens kønsfordeling og populationens kønsfordeling er relativt tættere på hinanden.

En anden forklaring er at drenge har en lavere sandsynlighed for at gennemføre undersøgelsen, og dermed vil kønsfordelingen blive skæv imod pigernes deltagelse i undersøgelsen. Dog vurderer vi, at forskellen er tilstrækkelig lille til at vi ikke vil efterkorrigere dette i undersøgelsen, og vi vurderer at kønsfordelingen i stikprøven er repræsentativ for kønsfordelingen i populationen.

Vi vil nu se på fordelingerne af tosprogede børn i undersøgelsen i henhold til populationen.

Tabel 14: Fordelingen af børnenes sproglige baggrund i henholdsvis stikprøven og populationen

	Stikprøven	Populationen
Personer med dansk oprindelse /et sprog	80,0%	88,3%
Indvandrere + efterkommere + uoplyst herkomst/to sprog i hjemmet	20,0%	11,7%

Anm. 1: Tallene er ikke direkte sammenlignelige. Se forklaring under tabellen.

Note 14: Egne optællinger på vægtede data. Populationsdata kommer fra DST.dk's tabel UDDAKT20

Vi gør opmærksom på at tallene i Tabel 14 ikke er direkte sammenlignelige. Det skyldes, at vi har spurgt om noget ikke-målbart for et offentligt register, nemlig hvor mange sprog der tales i hjemmet, mens Danmarks statistik bruger registerdata og deres definitioner af herkomst lavet ud fra fødselssted og statsborgerskab. Vi har registeret barnet som et barn med flere sprog i hjemmet, hvis barnet har angivet at de taler flere sprog i hjemmet, samt at det angivne sprog eksisterer. Vi har således valgt at sige, at de elever har angivet at de taler bornholmsk i hjemmet udover dansk kun taler et sprog i hjemmet. Vi har dermed valgt at se bort fra hvilket statsborgerskab eleven har, og vi bruger andre talte sprog i hjemmet som en proxy for andre kulturelle baggrunde hos eleverne. Med denne proxy vil vi kunne se, om der er kulturelle forskelle i hvordan eleven læser i sin fritid. Hos Danmarks Statistik er man registeret som etnisk dansk, hvis barnet er barn af en dansk statsborger, og barnet er født i Danmark. Det betyder, at barnet kan være registeret som etnisk dansker, men samtidigt have en forælder med en anden national oprindelse. Dermed kan et barn godt tale flere sprog i hjemmet, og stadigvæk være en person med dansk oprindelse. Siden vi tæller alle elever, der taler flere sprog i hjemmet, vil andelen af tosprogede elever være større i stikprøven end det Danmarks Statistik har registeret. Den talmæssige forskel opstår fordi de to optællingsmetoder er forskellige, men overordnet tror vi, at der ikke er nogen nævneværdig forskel imellem stikprøven og populationen, og vi vil derfor antage at vores stikprøve er repræsentativ for vores population. Yderligere tror vi, at den sproglige baggrund, som vi har spurgt ind til, vil give os et bedre billede af, hvordan den kulturelle baggrund kan påvirke hvordan fritidslæsningen er i hjemmet.

Vi vil nu se på årgangsfordelingerne og hvordan stikprøven adskiller sig fra populationen.

Tabel 15: Fordelingen af elever på årgangene i stikprøven og i populationen

	Stikprøven	Populationen
3. klasse	15,5%	19,8%
4. klasse	18,2%	20,2%
5. klasse	20,8%	19,8%
6. klasse	23,6%	20,1%
7. klasse	21,9%	20,0%

Anm. 2: Samlæste klasser er fjernet fra optællingen

Note 15: Egne optællinger på vægtet data. Populationsdata kommer fra DST.dk's tabel UDDAKT20

Som det kan ses i Tabel 15 passer andelen ikke helt til hinanden. Der er en lille overvægt af 6. og 7. klasser, mens der er relativt få 3. og 4. klasser i undersøgelsen. Andelen af 5. klasseelever er relativt tæt på den rette andel. Der er flere forklaringer på, at vi observerer disse forskelle imellem andelen i stikprøven og i populationen. Den første forklaring er, at vi har valgt ikke at indregne de samlæste klasser i vores sammenligning. Det skyldes at Danmarks Statistik kun har opdelt eleverne på hvilken klasseårgang, som elever *burde* være på. Men i virkeligheden er der samlæste klasser på skolerne, hvor eleverne fra flere klassetrin bliver lagt sammen til en klasse. Det betyder, at der kan ske en skævvridning, hvis tendensen til at samlæste klasser er højere på de mindre årgange. En anden forklaring er, at skoler har været mere villige til at deltage med de større klasser end de mindre klasser. Det betyder, at der vil være skoler, hvor der kun deltager ældre klasser. Vi har i vægtningen af eleverne inddraget dette element, at de deltagende klasser på skolen også skulle tale for alle de ikke-deltagende klasser på skolen. Dermed vil klasser på skoler, hvor det kun deltog ældre klasser, vægte mere, og det vil virke som om, at der var flere ældre klasser, og ikke yngre klasser, som de reelt skulle tale for. Umiddelbart vil vi vurdere, at betydningen af denne skævhed vil være minimal for vores undersøgelse. Vi vurderer, at det er en generel udfordring man vil have og dermed vil denne undersøgelse være det bedste bud på hvordan læsningen er i fritidslæsningen.

Vi vil nu se på fordelingen af eleverne på skoletype, hvorefter vi vil se på hvordan eleverne fordeler sig i stikprøven og populationen.

Tabel 16: Fordeling af elever på skoletype i stikprøven og i populationen

	Stikprøven	Populationen
Folkeskole	83,0%	82,5%
Privat- & friskoler	17,0%	17,5%

Note 16: Egne optællinger på vægtede data. Populationsdata kommer fra DST.dk's tabel UDDAKT20

Tabel 16 viser, at fordelingen af elever i folkeskolen og i privat- og friskoler er relativt ens i stikprøven og i populationen. Den opgjorte marginale forskel imellem de to andele skyldes, at der er udtrukket en marginalt større andel af folkeskoler i stikprøven end populationen angiver. Dog er den udtrækningsforskel så lille, at det ingen rolle har i vores analyse. Dermed vil vi vurdere, at den vægtede stikprøve i henhold til skoletype er repræsentativ.

Konklusionen på vores analyse af om den vægtede stikprøve er repræsentativ for vores population er, at vi vurderer den som repræsentativ. Dog skal man være lidt påpasselig med nogle sammenligninger, da klassefordelingen kan være lidt skæv. Dog er vurderingen at den eventuelle skævvridning er så lille, at vi sagtens sige stikprøven er repræsentativ for børn i alderen 9-13 år, og rapporten vil dermed kunne komme med generelle konklusioner om børns fritidslæsevaner.

Vægte

I vores stikprøve har vi valgt en stratificeret klyngeudvælgelsesstrategi, hvori vi siger, at alle skoler ikke har lige stor sandsynlighed for at blive udtrukket. Dette skyldes at antallet af skoler i hvert strata kan variere, selvom man udtrækker det samme antal skoler i flere stratum. Siden alle strata ikke har samme deltagelsesprocent og samme sandsynlighed for at blive udtrukket, har vi valgt at indføre vægte til at korrigere vores estimerede gennemsnit i den rette retning, således at vi får et retvisende resultat. Vi kommer til at lave forskellige vægte for eleverne og elevernes valg af titler og forfattere. Dog vil vægtene af elevernes valg af titler og forfattere afhænge af elevernes generelle vægte, og vi vil derfor gennemgå elevernes generelle vægte, for til sidst at tilføje hvad vi har gjort anderledes ved titel- og forfatter-vægtene.

Hver respondent får en vægt, som vi vil kalde for den totale vægt. Den totale vægt skal tolkes som hvor mange elever den enkelte elev taler for på landsplan. På den måde gør det ikke så meget, at der er nogle strata hvor alle elever har svaret, mens andre strata kun har få elever, der har svaret. Dette skyldes, at alle

elever vil udgøre en korrigeret andel af den samlede population. Hvis der bliver flere svar i et strata betyder det, at hver respondent inden for strataet vil vægte mindre. Den totale vægt kan deles op i flere vægte, som vi vil gennemgå efterfølgende.

Skoleudtrækningsvægten (SUV)

Skoleudtrækningsvægten (SUV) er vægten der justerer, at alle skoler ikke har lige stor sandsynlighed for at blive udtrukket. Denne vægt siger, hvor mange skoler den enkelte udtrukne skole skal tale for. Vægten udregnes som den inverse sandsynlighed for at blive udtrukket. Eksempelvis hvis en skole har 10% sandsynlighed for at blive udtrukket, bliver skoleudtrækningsvægten $\frac{1}{10\%} = \frac{1}{0,10} = 10$. Vi har udregnet sandsynligheden for at blive udtrukket som: $\frac{\text{Antallet af udtrukne skoler i strataet}}{\text{Antallet af skoler i strataet}}$

Det gør at skoleudtrækningsvægten

$$SUV = \frac{\text{Antallet af skoler i strataet}}{\text{antallet af udtrukket skoler i strataet}}$$

Den største skole udtrækningsvægt er 13,8, og det betyder, at en udtrukket skole fra dette strata taler for 13,8 skoler hver. Den laveste skoleudtrækningsvægt er 7, hvilket betyder at den udtrukne skole taler for 7 skoler.

Skoletilsagnsvægten (STV)

Skoletilsagnsvægten (STV) er vægten der justerer for hvor mange skoler, der skulle have deltaget i undersøgelsen fra strataet, og hvor mange skoler der har givet tilsagn om at deltage. Vi har valgt at inddrage denne vægt, da den vil give os en ide om hvor gennemførlig undersøgelsen er i de enkelte strata. Skoletilsagnsvægten vil yderligere justere de strata, hvor vi har fået flere skoler tilmeldt, end der var udtrukket fra strataet. Det skyldes, at de først adspurgte skoler ikke var hurtige nok til at tilmelde sig undersøgelsen, og at vi i mellemtiden havde fået en af erstatningsskolerne med i undersøgelsen. Skoletilsagnsvægten siger noget om, hvor nemt det har været at få det udvalgte antal skoler til at deltage fra hvert stratum. Desto større skoletilsagnsvægt, desto sværere har det været at få skolerne med.

$$STV = \frac{\text{Antallet af udtrukket skoler i strataet}}{\text{Antallet af skoler, der har givet tilsagn i deltagelse i undersøgelse i strataet}}$$

Den største skoletilsagnsvægt er 3, og det betyder, at en skole, der har givet tilsagn fra dette strata, vil skulle tale for 3 andre udtrukne skoler i strataet. Den mindste skoletilsagnsvægt er 0,5, hvilket betyder at de skoler, der har givet tilsagn i dette stratum, har været dobbelt så mange som de burde, og dermed taler de kun med halv kraft.

Skoledeltagelsesvægten (SDV)

Skoledeltagelsesvægten (SDV) er vægten der justerer for, hvor mange skoler der rent faktisk deltaget ud af dem, der har tilmeldt sig undersøgelsen. Det vil sige det inverse af deltagelsesprocenten fra hvert stratum, udregnet ud fra de tilmeldte skoler i strataet. Det kan med andre ord godt være skoler, der siger, at de deltager, men så alligevel ikke deltager. Det sker primært, fordi det er ledelsen, der har tilmeldt skolen til undersøgelsen, men det er lærerne, der har haft den endelige afgørelse om de vil deltage med klassen eller ej. Generelt gør det sig gældende, at der er 5 skoler med klasser, der har tilmeldt sig, men ikke har deltaget.

$$SDV = \frac{\text{Antallet af skoler, der har givet tilsagn i deltagelse i undersøgelse i strataet}}{\text{Antallet af skoler der har deltaget i undersøgelsen i strataet}}$$

Den største skoledeltagelsesvægt er 1,5, hvilket betyder at hver deltagende skole i strataet skal tale for 1,5 tilmeldte skoler i undersøgelsen. Den mindste skoledeltagelsesvægt er 1, hvilket betyder at alle de tilmeldte skoler i strataet deltager i undersøgelsen.

Elevudtrækningsvægten (EUV)

Elevudtrækningsvægten (EUV) er vægten der siger, hvor mange elever den udtrukne elev taler for. Det vil sige at elevudtrækningsvægten justerer den systematiske forskel i sandsynligheden for at blive udtrukket. Vægten udregnes som den inverse sandsynlighed for at blive udtrukket til projektet. I denne undersøgelse har vi valgt at sige, at alle eleverne fra de klasser, vi har kontakt til, er blevet udtrukket til undersøgelsen. Det vil sige at sandsynligheden for at blive udtrukket er 100%. Dermed betyder, at alle respondenters elevudtrækningsvægt er 1.

$$EUV = \frac{\text{Antallet af elever i klassen}}{\text{Antallet af udtrukket elever i klassen}}$$

Elevdeltagelsesvægten (EDV)

Elevdeltagelsesvægten (EDV) korrigerer for at antallet af deltagende elever i klassen ikke nødvendigvis svarer til det oplyste klasseantal. Det kan skyldes flere ting. Den primære årsag er dog elevens fravær på den pågældende testdag. Om det er sygefravær eller andet, er ligegyldigt for undersøgelsen. Vi er kun interesseret i om klassen udgør den andel, som de burde, i undersøgelsen. Elevdeltagelsesvægten korrigerer dette, og vi kan dermed sige, at vi får et robust estimat. Vægten udregnes som den inverse af deltagelsesprocenten i klassen. Det betyder, at jo lavere andel af eleverne i klassen, som har besvaret spørgeskemaet, desto mere vejer de deltagende elevers svar.

$$EDV = \frac{\textit{Antallet af udtrukne elever i klassen}}{\textit{Antallet af deltagende elever fra klassen}}$$

Den største elevdeltagelsesvægt er 4, hvilket betyder at de deltagende elever taler for 3 andre børn i klassen. Den laveste elevdeltagelsesvægt er 0,529, hvilket betyder at hver deltagende elev i klassen taler for 0,529 elev hver. Det kunne skyldes de tidligere omtalte deltagelsesprocenter fra klasserne, hvor der var nogle klasser med en deltagelsesprocent på over 100, hvis der er flere klasser der har brugt det samme link, som der kun var tiltænkt en bestemt klasse. Elevdeltagelsesvægten korrigerer dette, således at klassens svar vægtes som at alle i klassen har deltaget.

Klassedeltagelsesvægten (KLV)

Klassedeltagelsesvægten (KLV) er medregnet for at korrigere, at det ikke nødvendigvis er alle klasser på årgangen på skolen, som har deltaget i undersøgelsen. Ved at vægte de enkelte deltagende klasser op, således at de deltagende klassers besvarelser vejer som om at alle klasserne på årgangen har deltaget. Ved at bruge klasserne på de samme årgange på de samme skoler vil man have det bedste bud på, hvad de ikke-deltagende klasser ville have svaret i undersøgelsen, da det er tilfældigt hvilke elever der kommer i hvilke klasser. Besvarelserne i klasserne burde minde om hinanden, hvorved vi burde få et retvisende billede af elevernes brug af læsning i fritiden. Vægten udregnes som den inverse deltagelsesprocent for klasserne på årgangen fra skolerne.

$$KLV = \frac{\textit{Antallet af klasser på årgangen på skolen}}{\textit{Antallet af deltagende klasser på årgangen på skolen}}$$

Korrektionsvægten (Korrektion)

En udfordring ved klassesdeltagelsesvægten er, at der skal deltage mindst en klasse fra hver årgang på skolen, før at vægten virker perfekt. Det er desværre ikke altid tilfældet, således at nogle skoler kun deltager med enkelte årgange. Det betyder, at vi vil få en skævvridning, hvis vi ikke tager højde for denne udfordring. Vi har løst problematikken ved at danne en korrektionsvægt, som siger, at hvis en skole mangler besvarelser fra en hel årgang, vil de andre årgange træde til. Denne korrektionsvægt bliver udregnet som forholdet imellem antallet af klasser på årgangen på skolen, hvor ingen klasser fra årgangen deltager, i forhold til hvor mange klasser der deltager fra skolen i undersøgelsen.

Korrektion

$$= \frac{\text{Antallet af klasser på skolen fra årgange, hvor ingen klasser fra årgangen deltager}}{\text{Antallet af deltagende klasser på skolen}}$$

Korrektionsvægten kan antage forskellige værdier, og det er en stigende funktion af antallet af ikke-deltagende klasser på årgangen. Hvis mindst en klasse på hver årgang på skolen deltager og der dermed ikke mangler nogen årgange, så vil tælleren i korrektionsvægten være 0, og korrektionsvægten vil dermed også være 0. Men i takt med at antallet af årgange, der ikke deltager stiger, stiger korrektionsvægten. Vi har valgt denne løsning, da elever på samme skole må minde om hinanden i form af social og kulturel baggrund, da eleverne kommer fra samme skoledistrikt. Det gælder også selvom de deltagende elever er ældre eller yngre end dem der mangler at deltage. Det er typisk, at det er de ældre klasser, der deltager i undersøgelsen, og dermed vil de veje mere i undersøgelsen. For at lave den samlede vægt for klassesdeltagelsen, har vi valgt at lægge de to vægte for klassesdeltagelsesvægten og korrektion sammen. Fortolkningen af denne vægt vil være, hvor mange klasser hver deltagende klasse på skolen deltager for. Den minimale vægt for denne klassevægt er 1, hvilket betyder at alle klasser på skolen har deltaget, og hver klasse kun taler for sig selv. Den maksimale vægt for denne klassevægt er 11, hvilket betyder at den deltagende klasse taler for yderligere 10 klasser på skolen. Det er ligegyldigt om alle de ikke-deltagende klasser er på samme årgang, som den deltagende klasse er. Det skyldes, at de deltagende elever på skolen vil være den bedste proxy for de ikke-deltagende elever på skolen.

Afslutningsvægten

Afslutningsvægten sikrer, at vi kun ser på respondenter, der har fuldført spørgeskemaet. Det sikrer blandt andet, at udregningerne ikke medtager de frafald, som sker i løbet af besvarelsen af spørgeskemaet.

Afslutningsvægten har værdien 1, hvis respondenter har fuldført spørgeskemaet, og ellers 0. Det sidste sker ligegyldigt om respondenter ikke har påbegyndt spørgeskemaundersøgelsen eller stopper undervejs.

Totalvægte (TWGT)

Til sidst har vi totalvægten (TWGT), som samler alle vægtene på følgende måde:

$$Twgt = SUV * STV * SDV * EUV * EDV * (KLV + Korrektion)$$

Totalvægten siger, hvor mange elever den givne elev, der har deltaget i undersøgelsen, taler for i forhold til hele populationen. Her er populationen det samlede antal af elever i 3.-7. klasse i enten folkeskolen eller privat- eller friskoler. Vi har fra Danmarks Statistiks tabel UDDAKT20, at der er 334.418 elever i vores population. Vi har optalt det vægtede antal af respondenter til at være 327.098 elever. Det svarer til en fejlmargen på -2,2%. Det er naturligt, at vi har estimeret et samlet antal af elever, der er mindre end populationsantallet. Det skyldes, at vi har lavet vægtningen ud fra de strata, vi har udtrukket fra. Da nogle strata havde for få skoler, udtrak vi ikke nogle skoler derfra. Antallet af elever i disse strata svarer meget godt til 7.000 elever, hvilket svarer til fejlmarginen. Vi ved fra tidligere, at vores stikprøve for de udtrukne var repræsentativ for hele populationen, og dermed vil vi have et retvisende bud, selvom vi har en fejlmargen på 7.000.

Igen må vi understrege, at det er vigtigt at bruge totalvægten, når man skal finde forskellige estimater i undersøgelsen. Dette skyldes stikprøveudtrækningsstrategien, hvor alle skoler og elever ikke har samme sandsynlighed for at blive udvalgt. Når man bruger totalvægten vil man få et unbiased estimat fra stikprøvedesignet.

Replikationsvægte

Brugen af replikationsvægte er en metode, hvorved man ud fra en stikprøveundersøgelse kan lave "flere" undersøgelser ud fra det samme datasæt. Dette sker ved at ændre respondenternes vægte på en systematisk måde, således at man kan få 111 datasæt, hvor hvert datasæt kan få sin unikke kombination af totalvægte. Når man bruger replikationsvægte får man et bedre estimat for standardfejlen af ens gennemsnit, hvilket skyldes at der fjernes noget af usikkerheden fra vægtene, som kommer fra vægtkonstruktionen.

I undersøgelsen "Børns læsning 2017" har vi valgt at bruge JackKnife replikationsvægte, da de er nemme at skabe, og da de giver et tilstrækkeligt præcist estimat på standardfejlen til denne type af undersøgelse. Hver JackKnife replikationsvægt består af de samme vægte som totalvægtene, med undtagelse af to skoler, som vil have andre vægte. Disse to skoler er et par, og alle skoler er sat sammen i par. I hver replikationsvægt vil parret, hvor vi har andre vægte end totalvægtene, få andre vægte. Den ene skole vil have fået værdien 0, mens den anden skole vil få fordoblet sin vægt. I den næste replikationsvægt vil det være omvendt. I "Børns læsning 2017" har vi valgt at lave to replikationsvægte for hvert skolepar. Siden vi har 110 skoler med i "Børns læsning 2017" har vi valgt at skabe 55 skolepar med to replikationsvægte hver, således at vi samlet har 110 replikationsvægte. Hvert par er skabt tilfældigt, da vi blot ønsker at skabe replikationsvægtene, således at vi kan få "flere" stikprøver ud fra vores undersøgelse. Når vægtene for en skole varierer på tværs af replikationsvægtene, vil summen af de totale vægte i hver enkelt replikationsvægt variere. Det har dog ingen betydning for estimerne, da udregningerne foregår via gennemsnitsberegningerne, hvor man dividerer med summen af totalvægtene i hver replikationsvægt.

Skoleparrene er skabt via Undervisningsministeriets institutionsnumre, sorteret stigende, hvor vi har taget de første to skoler og dannet et par, og efterfølgende taget de næste to skoler til næste par. Denne procedure er forsat indtil alle deltagende skoler var i et par. Vi har dermed valgt at lade tilfældighederne bestemme hvilke skoler, der skulle danne et par. Hvis der skulle være en systematik bag Undervisningsministeriets institutionsnummer, har vi ikke haft kendskab til denne systematik, og dermed kan vi ikke tage højde for det. Fordelen ved at lade tilfældighederne råde er, at vi får robusthedstestet vores estimer. Grunden til at vi får robusthedstestet vores estimer ved denne metode er, at summen af vægtene i replikationsvægtene varierer på tværs af replikationsvægtene og dermed vil forskellige skoler vægte relativt forskelligt. Årsagen til dette er at vores skoler i skolepar kan komme fra forskellige strata, og dermed vil vi have vidt forskellige skoler i vores skolepar. Dermed vil vi kunne have et par, hvor den første skole taler for 40 skoler, mens den anden skole i parret kun taler for 10 skoler. Når vi i JackKnife proceduren fordobler den enes skoles vægt, og den anden skoles vægt bliver 0 i en replikationsvægt, betyder det at der vil være en forskel på 60 i totalvægten. Det vil kunne ændre vores gennemsnit betragteligt, og dermed vil vi få et mere robust estimat for standardfejlen af vores gennemsnit.

Der er to andre mulige strategier for dannelsen af parrene til replikationsvægtene. Den ene potentielle strategi er, at vi sorterer skolerne ud fra hvilket strata de tilhører, og derefter danner parrene ud fra denne liste. Dette vil gøre, at vi i langt de fleste tilfælde vil have relative ens skoler i hvert skolepar. Grunden til at det bliver ens er, at skolerne inden for de forskellige strata vil være forholdsvis ens, da vi har meget specifikke strata skabt ud fra hvordan skolerne er. Den anden strategi er at vi kunne havde sorteret

skolerne efter den totale vægt på skolen. Det vil give, at summen i de totale vægte i hver replikationsvægt vil være relativt ens. Hvis vi havde valgt en af de to andre strategier for pardannelse, ville vi ikke i samme grad have fået robusthedstjekket vores estimater, og dermed ville vores estimat for standardfejlen være mere tvivlsomt, end med vores valgte strategi.

Vægte vedrørende titler og forfattere

Alle elever har haft mulighed for at angive tre forfattere og titler, som de har læst i deres fritid i den seneste periode. For at sikre at der ikke kommer en skæv analyse ud af elevernes svar, har vi valgt at ensrette elevernes angivne svar, samt konstruere nye vægte ud fra hvor mange mulige svar, som de har kunnet angive. Vedrørende ensretningen af titler og forfattere, se længere ned i rapporten. I praksis er vægtene vedrørende elevernes oplysninger om forfattere og titler lavet på den måde, at alle elevernes vægte er blevet reduceret med en tredjedel, ligegyldigt om de har angivet tre titler eller nul titler. Det skyldes, at vi i stedet vurderer, hvor mange gange er en bog nævnt af det mulige antal på $3 \cdot 8721 = 26163$. Alle de steder hvor eleverne ikke har angivet nogen titler, er det blevet kodet som et ugyldigt svar. Dermed har alle elever "angivet" tre muligheder, og dermed vil korrektionen af vægtene på titlerne gøre det væsentlig nemmere, når vi udregner andelen af brugen. For at kunne lave udregninger har vi konstrueret et særskilt datasæt, hvorfra vi har udregnet top ti-listerne på de enkelte grupperinger. Titel- og forfattervægtene sikrer, at vi ikke får et skævt resultat.

Spørgeskemaets inspirationskilder

Spørgeskemaet er blevet inspireret af blandt andet PIRLS 2016-undersøgelsen, hvor vi har taget spørgsmålene om elevernes selvvaluerede dygtighed i læsning. Siden vi ikke har nogle kompetencetest i undersøgelsen vil man kunne bruge PIRLS-undersøgelsen i en argumentation for en sammenhæng imellem elevernes dygtighed, selvvaluerede dygtighed, og deres læsning i fritiden. PIRLS-spørgsmålene bliver således et bindeled imellem de to undersøgelser. PIRLS-undersøgelsen har givet inspiration til spørgsmål i denne undersøgelse, hvilket vil sige, at vi har taget tematikker inden for feltet, og indtænkt hvordan PIRLS-spørgeskemaet er konstrueret. Yderligere har vi taget udgangspunkt i spørgeskemaet fra Stine Reinholdt Hansens ph.d.-afhandling fra 2014, hvori hun undersøgte elevers fritidslæsning i 2010. Vi har valgt at bruge denne undersøgelse som en benchmark for resultaterne i denne undersøgelse. Det vil sige, at vi har gentaget flere af spørgsmålene fra 2010-undersøgelsen i 2017-undersøgelsen.

Rensning af data

Vi har valgt kun at se på fuldførte besvarelser, da vi således er sikre på at børnene har besvaret alle spørgsmålene, og dermed vil vi nemt kunne skabe indeks på spørgsmålene. Yderligere giver det den fordel, at respondentantallet er konstant. Spørgeskemaet er konstrueret således, at næsten alle svar, som børnene skulle give, har været i et multiple choice-format. Det vil sige, at eleverne i begrænset omfang skulle skrive, og dermed har vi minimeret sandsynligheden for fejlindtastninger. 4 steder i spørgeskemaet har vi givet eleverne mulighed for at skrive deres svar. Det drejer sig om, hvilket andet sprog de taler i hjemmet, hvor elevernes finder inspiration til læsning, og hvor de ellers finder inspiration til læsning på internettet, samt eleverne skulle angive titlen og forfatteren på de bøger, som de har læst inden for den seneste måned. Eleverne har haft tre muligheder for at angive titler og forfattere. Selve behandlingen af de angivne titler og forfattere er forklaret i afsnittet neden for. Vedrørende de tre andre kvalitative svarmuligheder har vi ikke lavet en ensretning af elevernes svar. Vi har dog kigget på elevernes angivne svar på hvilket andet sprog de taler i hjemmet, og kategoriseret svarene i sproggrupper.

Vedrørende hvilket sprog børnene taler udover dansk i hjemmet, har vi valgt at konstruere en sproglig baggrundsvariabel ud fra vores valgmulighedsliste og elevernes svar. Vi har valgt at lave følgende sproggrupper: Vestlige sprog, slaviske sprog, mellemøstlige sprog, fjernøstlige sprog, afrikanske sprog, samt nordiske sprog. Vestlige sprog omfatter engelsk, tysk, fransk, spansk, italiensk, flamsk, portugisisk, samt lignende sprog. Slaviske sprog omfatter polsk, tjekkisk, sprog på Balkan, russisk, og baltiske sprog. Til mellemøstlige sprog tilhører tyrkisk, de nordafrikanske sprog, israelsk, arabiske sprog, afghansk, samt pakistansk. Til fjernøstlige sprog er sprog øst for den indisk-pakistanske grænse. Det vil sige hindu, kinesisk, japansk, koreansk, thai, indonesisk, vietnamesisk, filippinsk, malaysisk, med videre. Afrikanske sprog er sprog tilhørende lande syd for Sahara. Til Nordiske sprog hører norsk, svensk, finsk, islandsk, grønlandsk og færøsk. Hvis det angivne sprog ikke kunne placeres i en af kategorierne, så betegnes det som ugyldigt, og dermed vil eleven kun tælle som etsproget. Et eksempel på et ugyldigt sprog er bornholmsk, som vi har betegnet som dansk. Variablen tosproget er konstrueret således at man er tosproget, hvis man tilhører mindst en af kategorierne af sproggrupper. Det vil sige, hvis en elev både taler dansk, russisk og engelsk, vil eleven tilhøre både den vestlige og slaviske sprogbaggrund, men eleven vil kun betegnes som tosproget, og ikke som tresproget. Vi har valgt kun at tage kategorierne etsprogede og tosprogede med, da det vil være mere simpelt, samt at gruppen af tresprogede elever kun består af 31 børn ud 8721 mulige.

Angående de to andre muligheder for at angive kvalitative svar, har vi ikke kategoriseret eller ensrettet svarene. En af grundene til at vi ikke har ensrettet dem er, at svarene vil give os marginal mere viden om elevernes inspirationskilder. Enten fordi børnene har skrevet noget, som de har kunnet vælge på vores

liste, eller fordi de har skrevet noget, som er unikt for dem, såsom "4chan.com". Vi vil bruge børnenes angivne svar som en inspirationskilde til fremtidige undersøgelser, således at vi kan få mere præcis viden i andre undersøgelser.

Pilottestning af spørgeskemaet

Vi har pilottestet et oprindeligt spørgeskema i december 2016 på en privatskole og en folkeskole med klassetrinene 3., 4. og 7. klasse. Målene med pilottesten var at teste om spørgsmålene fungerede, hvor langt tid besvarelsen tog, og hvordan selve testafholdelsen fungerede. Testafholdelsen foregik således, at vi havde konstrueret et "unik" spørgeskema til hver klasse, således at besvarelser fra hver klasse kom i sin egen surveydatabase i SurveyXact. Dette foregik uden problemer, hvilket gjorde, at vi gentog proceduren i selve hovedundersøgelsen. Selve testafholdelsen var for lang i forhold til vores mål, hvilket gjorde, at vi fjernede nogle spørgsmål (fx angående alder og klassetrin), da vi allerede havde en ide om disse forhold fra centralt hold af. Da hver klasse fik sit unikke spørgeskemalink, ved vi hvilket klassetrin der har besvaret hvilket link. Det betyder, at vi kender elevernes omtrentlige alder.

Til sidst var der problematikken om hvorvidt selve spørgsmålene fungerede. Vi tog udgangspunkt i faktoranalyser, hvor vi så at nogle spørgsmål ikke fungerede i henhold til vores andre spørgsmål i spørgsmålsbatteriet. Hvis et spørgsmål ikke gav os yderligere, relevant viden i forhold til de resterende spørgsmål, betød det at spørgsmålet blev taget ud af spørgeskemaet. Alt i alt tjente pilotundersøgelsen sit formål. Spørgeskemaet blev skarpere og testafholdelsen blev testet og fundet fornuftig.

Svarrækkefølgen

Vi har i undersøgelsen valgt at konstruere et forsøg med svarrækkefølge effekter. Det vil sige, at vi ville undersøge om det har en betydning for hvordan eleverne i 3. - 7. klasse svarer, om de møder "Helt enige" svarmuligheder først eller om de møder "Helt uenige" svarmuligheder først. Da vi har haft 8721 elever fordelt i mange klasser, har vi valgt at lave et forsøg, hvor alle har fået det samme spørgeskema, men svarrækkefølgen har enten været i form af uenig til enig, eller enig til uenig. Det er fuldstændigt tilfældigt hvilken version af spørgeskemaet klasserne har fået. Det betyder, at Center for Anvendt Skoleforskning kan lave forskning på svarrækkefølgerne i folkeskolen uden at det går ud over kvaliteten af denne undersøgelse. Yderligere vil denne opdeling muligvis give et mere retvisende resultat, da eventuel bias grundet svarmulighedernes rækkefølge ikke bliver en faktor i denne undersøgelse.

Ensretning af forfattere og titler

Eleverne har i spørgeskemaet fået et spørgsmål hvor de kunne angive de sidste tre bøger, som de har læst i fritiden. Spørgsmålet er præciseret til at det drejer sig om ikke-lektie-relateret læsning. Børnene har således skulle skrive titel og forfatter, hvis de kan huske begge. Siden vi her arbejder med selvskrivne titler og forfattere, vil der være mange forskellige måder at skrive titler og forfatterne på. For at vi kan arbejde videre med titlerne og forfatterne statistisk, har vi valgt at få alle titlerne og forfatterne ensrettet. På den måde vil en bestemt titel og en bestemt forfatter kun optræde på en bestemt måde og dermed er det nemmere at lave statistik på.

Hver elev har haft mulighed for at skrive 3 titler hver, og med 8.721 deltagere i undersøgelsen vil det give os en maksimal titelliste på 26.163 bøger. En stor del af disse 26.163 muligheder er ikke blevet brugt således at svarfelterne var blanke. Vi har valgt at sige alle blanke svar er ugyldige svar, således at eventuelle beregninger bygger på alle de mulige svar, som har kunnet været givet, og ikke på hvor mange elever der reelt set læser det. Det ændrer dog intet i rangeringen af bøger, som børnene læser. Dermed vil vores top ti-liste være i den samme rækkefølge, ligegyldigt om det var opgjort på individer eller antal nævnte gange. Ud fra de 26.163 besvarelser er der 11.688 forskellige kombinationer af titler og forfattere, der skulle ensrettes.

Ensretningen er foregået ved, at vi har haft en studentermedhjælper til at ensrette data så meget som muligt. Det vil sige at studentermedhjælperen har siddet og skrevet alle svar på samme bog, således at de stod stavet på samme måde. Eksempler kunne være at "Harry Pottter" og "Hary Poter" begge blev ensrettet til "Harry Potter". Et vigtigt element er, vi kun kan skrive de svar som børnene har givet os. Vi ved således ikke, når eleverne har skrevet "Harry Potter", hvilken en af "Harry Potter"-bøgerne der er tale om. Vi har i de tilfælde blot skrevet "Harry Potter" for at sikre en konsistens, og vi kan dermed udregne hvor mange der læser hvilke titler. For at bedre kunne sammenligne grupper af bøger, har vi valgt at lave en variabel, der hedder serier. Denne variable indikerer hvis bogen indgår i en serie, således at vi kan observere hvilke serier der er mest populære og hvor populært serieformatet i sig selv er. Fordelen er også, at vi ikke er afhængige af om børnene har præciseret hvilken bog i serien, de har læst, da alle angivelser tæller med i samme serie. Det vil sige at "Harry Potter" og "Harry Potter og de Visers Sten" begge tæller med i serie-opgørelsen, mens de vil tælle i hver sin kategori i titelopgørelsen.

Metoden hvormed ensretningen er foregået, er at vi har startet med at ensrette titlerne, således at de i høj grad er stavet på samme måde. Dette er sket med udgangspunkt i børnenes angivelser, og hvor vi med en

stor sikkerhed har kunnet identificere den titel, som børnene mente. I langt de fleste tilfælde har der ikke været tvivl om titlen, og det har dermed været relativt nemt at ensrette titlerne. Oftest gælder det, at børnene har haft nemmere ved at huske titlerne end forfatterne, og det gør at mange titler ikke umiddelbart var koblet til en forfatter. Vi har løst dette ved at sortere alle titler, således at alle ensrettede titler stod under hinanden. Herfra kunne forfatternavnene ensrettes, og vi har kunnet udfylde de steder, hvor der manglede en forfatter. For at lave en ekstra kontrol har vi sorteret efter listen med forfattere for at se om der var nogle titler, der skulle tilrettes i forhold til hvordan den korrekte titel staves. Når alle listerne med kombinationen af titler og forfattere er komplette, har vi undersøgt hvilke forfattere, der står bag de titler, som manglede en forfatter. Hvis der har været flere forfattere til den samme bog, har vi valgt at skrive den første forfatter på listen, således at vi kan lave statistiske beregninger på de oplyste titler og forfattere. Hvis et barn har oplyst 3 bøger i den samme serie, vil det tælle med 3 gange og ikke som 1. Fortolkningen af top ti-listerne er ikke direkte hvor mange der læser disse bøger, men hvor mange gange de er nævnt. Der er dog en stor sammenhæng, da hvert barn maksimalt kun kunne angive 3 titler og forfattere. Top ti listerne kan ikke bruges direkte til at se hvor mange der læser en given titel. Det skyldes at børnene kun kunne angive 3 titler og forfattere. Dermed kan de have læst andre og flere titler. Listerne giver en god indikator på hvad der er mest udbredt blandt børn i undersøgelsesperioden.

Faktoranalyser

Spørgsmålsbatterierne i "Børns læsning 2017"-spørgeskemaundersøgelsen er konstrueret således, at de burde måle bagvedliggende fænomener. Det vil sige, at vi har stillet spørgsmålene så klart og tætstående som muligt, således at det er den samme bagvedliggende faktor, der slår igennem i måden, hvorpå barnet svarer på spørgsmålene. For at kunne lave mere avanceret analyser senere hen har vi foretaget faktoranalyser i de enkelte spørgsmålsbatterier for at identificere de bagvedliggende faktorer. Et spørgsmålsbatteri er en samling af spørgsmål angående det samme tema. Et eksempel på et tema kunne være en række spørgsmål angående, hvor gode eleverne synes, at de er til at læse. Vi har i alt fundet 16 faktorer i datasættet, og de er alle lavet om til indeks. Udgangspunktet for vores rotation, hvis der har været flere dimensioner i spørgsmålsbatteriet, har været Varimax-rotation. Hvis vi synes at denne rotation ikke gav et klart nok billede af dimensionerne, har vi lavede Promax-rotation som supplement. Hvis vi har lavet Promax-rotation, har resultaterne fra Promax-rotationen været udgangspunkt for indekskonstruktionen. Varimax-rotationen gør, at dimensionerne i faktoranalysen optræder uafhængige af hinanden.

Her har vi resultatet af spørgsmålsbatteri 3, hvor vi har spurgt til, hvad eleverne tænker om at læse.

Tabel 17: Faktoranalyse for spørgsmål i spørgsmålsbatteri 3

Pattern Matrix^a

	Component	
	1	2
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg læser kun, hvis jeg er nødt til det	0,835	0,091
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg er nødt til at kunne læse godt af hensyn til min fremtid	0,160	0,768
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg synes, læsning er cool	-0,664	0,209
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg læser mest, når de voksne siger, jeg skal	0,839	0,184
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Mine forældre går ikke op i, om jeg læser	0,034	-0,457
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jo mere jeg læser, jo bedre bliver jeg	-0,046	0,742
3. Hvad tænker du om at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg synes, læsning er for nørder	0,462	-0,384

Note 17: Promax rotation. N=8721

I ovenstående Tabel 17 har vi markeret med gult hvilke spørgsmål, der tilhører hvilken dimension. Vi kan se, at der er 4 spørgsmål i den første kolonne, som udgør en dimension, mens vi kan se at der er 3 spørgsmål i den anden kolonne som udgør en anden dimension. Vi har valgt at døbe den første dimension: "Læselyst", og den anden dimension er døbt "Læsning fra et nytte- og præstationsperspektiv samt forældrenes engagement i forhold til deres barns læsning".

Ved en reliabilitetsanalyse kan vi se, at dimensionen "Læselysten" har en relativ høj Cronbachs alpha.

Tabel 18: Cronbachs alpha for indekset "Læselysten"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,702	4

Note 18: N=8721

Med en Cronbach's alpha på 0,702 er reliabiliteten på indekset høj, og vi vil derfor fastholde indekset "Læselyst" i vores datasæt.

For indekset "Læsning fra et nytte- og præstationsperspektiv samt forældrenes engagement i forhold til deres barns læsning" får vi en lavere Cronbach's alpha end ved "Læselyst" dimensionen.

Tabel 19: Cronbach's alpha for indekset "Læsning fra et nytte- og præstationsperspektiv samt forældrenes engagement i forhold til deres barns læsning"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,417	3

Note 19: N=8721

Med Cronbach's alpha på 0,417 er reliabiliteten på indekset meget lav, og reelt set burde vi smide indekset ud. Selve indekset består af spørgsmålene: "Jeg er nødt til at kunne læse godt af hensyn til min fremtid", "Mine forældre går ikke op i, om jeg læser" og "Jo mere jeg læser, jo bedre bliver jeg". Når man ser på spørgsmålene, så kredser de ikke helt om det samme klare fænomen, og dermed har indekset en lav reliabilitet. Vi har valgt at skabe indekset alligevel, da det vil give mulighed for perspektivering og mulighed for flere analyser. Yderligere vil indekset kunne blive brugt som en indikator.

Her ser vi på spørgsmålsbatteri 4, hvor vi har spurgt eleverne om, hvor gode de er til at læse.

Tabel 20: Faktoranalysen for spørgsmål i spørgsmålsbatteri 4.

Component Matrix^a		Component 1
4. Hvor god er du til at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg klarer mig som regel godt i læsning		-0,757
4. Hvor god er du til at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Det er nemt for mig at læse		-0,806
4. Hvor god er du til at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg har svært ved at læse historier med svære ord		0,638
4. Hvor god er du til at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg læser ikke så godt som andre i min klasse		0,764
4. Hvor god er du til at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Læsning er sværere for mig end noget andet fag		0,757
4. Hvor god er du til at læse? Fortæl, hvor enig du er i hvert udsagn (Sæt et X i hver linje) - Jeg er bare ikke god til at læse		0,789

Note 20: N=8721

I ovenstående Tabel 20 kan vi se at alle spørgsmålene udgør en faktor. Vi har valgt at døbe faktoren "Læsekompetence". Det er ikke så underligt, at alle spørgsmålene går i en dimension, da spørgsmålene kommer fra PIRLS undersøgelsen, som har specialiseret sig i at konstruere spørgsmål, der tilhører en enkelt dimension. Dimensionen er sikret således, at den gør sig gældende på tværs af lande og årgange, dermed vil dimensionen også træde frem i vores faktoranalyse.

Ved en reliabilitetstest af dimensionen "Læsekompetence" finder vi relativ høj Cronbach's alpha.

Tabel 21: Cronbach's Alpha for indekset "Læsekompetence"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,842	6

Note 21: N=8721

Vi kan se, at indekset "Læsekompetence" har en høj Cronbach's alpha, og vi vil derfor inddrage det i vores datasæt. Via faktoren "Læsekompetence" kan man drage koblinger imellem PIRLS undersøgelsen og "Børns læsning 2017"-undersøgelsen.

Her ser vi på spørgsmålsbatteri 8, hvor vi har spurgt børnene om, hvor ofte de gør forskellige ting uden for skolen.

Tabel 22: Faktoranalyse af spørgsmål i spørgsmålsbatteri 8

Component Matrix^a

	Component 1
8. Hvor ofte gør du disse ting uden for skolen? (Sæt et X i hver linje) - Jeg læser højt for nogen derhjemme	0,717
8. Hvor ofte gør du disse ting uden for skolen? (Sæt et X i hver linje) - Jeg lytter til, at nogen læser højt for mig derhjemme	0,683
8. Hvor ofte gør du disse ting uden for skolen? (Sæt et X i hver linje) - Jeg snakker med mine venner om det, jeg læser	0,647
8. Hvor ofte gør du disse ting uden for skolen? (Sæt et X i hver linje) - Jeg snakker med min familie om det, jeg læser	0,768

Note 22: N=8721

I ovenstående Tabel 22 kan vi se at alle spørgsmålene udgør en faktor. Vi har valgt at døbe dimensionen "Læserelationen". Vi har selv skabt disse spørgsmål, og vi er ganske tilfredse med, at de går ind på en enkelt dimension.

Ved en reliabilitetstest af dimensionen "Læserelationen" finder vi en fin Cronbach's alpha.

Tabel 23: Cronbach's alpha for indekset "Læserelationen"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,661	4

Note 23: N=8721

Vi kan se at indekset "Læserelationen" har en Cronbach's alpha på 0,661, hvilket vi betragter som acceptabelt, og vi vil derfor beholde indekset i datasættet.

Vi vil nu se på faktoranalysen af spørgsmålsbatteri 9a, hvor vi har spurgt børnene, hvor ofte de læser analoge tekster i fritiden.

Tabel 24: Faktoranalyse af spørgsmål i spørgsmålsbatteri 9a

Rotated Component Matrix^a

	Component	
	1	2
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser tegneserier	0,550	0,071
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser historier eller romaner	0,216	0,409
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser bøger, som forklarer ting (fx sport, idoler, dyr, mode)	0,431	0,325

9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser blade og magasiner (fx Vi unge, Illustreret Videnskab, Goal)	0,140	0,684
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser aviser	0,349	0,481
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser religiøse tekster (fx Bibelen eller Koranen)	0,676	0,020
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser instruktioner eller manualer (fx til spil eller legetøj)	0,497	0,237
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser undertekster på TV	-0,107	0,735
9a. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser digte	0,619	0,098

Note 24: Varimax rotation. N=8721.

Vi kan se i Tabel 24 at spørgsmålene fordeler sig i 2 dimensioner, som vi desværre ikke har nogle gode navne til. Variationen imellem de enkelte spørgsmål i de enkelte dimensioner er for stor til at vi kan sige, at de dækker et entydigt tema. Vi har bare valgt at døbe dimensionerne faktor_9_a_1 og faktor_9_a_2.

I en reliabilitetsanalyse får vi at indekset faktor_9_a_1 har følgende Cronbach's alpha

Tabel 25: Cronbach's alpha for indekset faktor_9_a_1

Reliability Statistics	
Cronbach's Alpha	N of Items
0,529	5

Note 25: N=8721

Vi kan se at indekset faktor_9_a_1 har en Cronbach's alpha på 0,529, hvilket er lavt. Vi har dog valgt at fastholde indekset i datasættet, men er meget påpasselige med at bruge det, da vi ikke er helt sikre på, hvad det udtrykker.

I en reliabilitetsanalyse får vi, at indekset faktor_9_a_2 har et Cronbach's alpha på 0,456.

Tabel 26: Cronbach's alpha for indekset faktor_9_a_2

Reliability Statistics	
Cronbach's Alpha	N of Items
0,456	4

Note 26: N=8721

Vi kan se at indekset faktor_9_a_2 har en Cronbach's alpha på 0,456, hvilket er for lavt. Vi har dog valgt at fastholde indekset i datasættet, men vi vil ikke bruge indekset. Det skyldes, at tilliden til indekset er lavt, samt at vi ikke helt ved hvad indekset giver udtryk for. Vi har valgt at fastholde det, da indekset kan give os nogle indikatorer på, hvor vi muligvis kan finde noget interessant blandt børnene.

Vi vil nu se på faktoranalysen af spørgsmålsbatteri 9b, hvor vi har spurgt til hvor ofte de læser digitale tekster i deres fritid.

Tabel 27: Faktoranalyse af spørgsmål i spørgsmålsbatteri 9b

Rotated Component Matrix^a

	Component	
	1	2
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser netaviser	0,497	0,152
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser beskeder på telefon, computer, tablet (fx Messenger, SnapChat, SMS, Twitter, Mail)	0,674	-0,144
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser web-sider (fx bold.dk, YouTube, Google, Facebook, Wikipedia)	0,724	-0,132
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser sangtekster (Lyrics)	0,663	0,167
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser blogs	0,618	0,322
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg læser e-bøger (Fx fra eReolen.dk, Superbog.dk, Frilæsning.dk, Mofibo)	0,108	0,818
9b. Hvor ofte læser du disse ting uden for skole (ikke lektier)? (Sæt et X i hver linje) - Jeg hører/læser lydbøger	-0,006	0,837

Note 27: Variamax rotation. N=8721.

Vi kan i Tabel 27 se at faktoranalysen af spørgsmålsbatteri 9b siger, at der skal være 2 dimensioner. Vi har valgt at døbe den første dimension for "net-afhængig læsning", mens den anden dimension er døbt "ikke-net-afhængig læsning". Her er sondringen, at e-bøger og lydbøger kan hentes ned, og læses og høres uden, at der er internet til stede.

I en reliabilitetsanalyse får vi at faktoren net-afhængig læsning har en Cronbach's alpha på 0,646.

Tabel 28: Cronbachs alpha for indekset net-afhængig læsning

Reliability Statistics

Cronbach's Alpha	N of Items
0,646	5

Note 28: N=8721

Vi kan se at indekset "net-afhængig læsning" har et Cronbach's alpha på 0,646, hvilket er acceptabelt. Vi har derfor valgt at fastholde indekset i datasættet.

I en reliabilitetstest for indekset "ikke-net-afhængig læsning" er Cronbachs alpha på 0,641.

Tabel 29: Cronbachs alpha for indekset ikke-net-afhængig læsning

Reliability Statistics

Cronbach's Alpha	N of Items
0,641	2

Note 29: N=8721.

Med en Cronbachs alpha på 0,641 har indekset "ikke-net-afhængig læsning" en acceptabel værdi. Det betyder, at vi har tillid til indekset, og at det måler, hvad vi forventer. Det vil sige, at jo mere børn læser e-

bog og hører lydbøger, desto højere scorer de i indekset. Med den acceptable værdi af Cronbach's alpha for indekset, har vi valgt at fastholde indekset i datasættet.

Vi vil nu se på en faktoranalyse i spørgsmålsbatteri 13c, hvor vi har spurgt eleverne om, hvad de synes om skolebibliotekerne på skolen. Her skal vi være opmærksomme på, at spørgsmålet kun er besvaret af elever, der har sagt, at de har et skolebibliotek på skolen. Der er i alt 8721-7835=886 elever, som har sagt, at de ikke har et skolebibliotek.

Tabel 30: Faktoranalyse af spørgsmål i spørgsmålsbatteriet 13c

	Rotated Component Matrix ^a	
	Component 1	Component 2
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - Jeg kan spørge dem om alting	0,766	-0,187
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - De giver mig gode tips til bøger	0,750	-0,255
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - De lytter	0,845	-0,182
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - De er venlige	0,787	-0,154
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - De er ikke så nemme at snakke med	-0,368	0,592
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - Jeg snakker aldrig med skolebibliotekerne	-0,141	0,850
13c. Hvad synes du om skolebibliotekerne på skolebiblioteket? - Jeg ser dem næsten aldrig	-0,149	0,822

Note 30: Varimax rotation. N=7835.

I Tabel 30 kan man se at faktoranalysen har givet, at der i spørgsmålsbatteriet er to dimensioner. Den første dimension har vi døbt "skolebibliotekarernes imødekommenhed", og den anden har vi kaldt "skolebibliotekarernes tilgængelighed".

I en reliabilitetsanalyse af faktoren "Skolebibliotekarernes imødekommenhed" er der Cronbach's alpha på 0,827.

Tabel 31: Cronbach's alpha for indekset "Skolebibliotekarernes imødekommenhed"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,827	4

Note 31: N=7835

Med en Cronbach's alpha for indekset "Skolebibliotekarernes imødekommenhed" på 0,827, er der en høj tillid til indekset måler det samme, og det måler hvor imødekomne skolebibliotekerne på skolen er. Vi har derfor valgt at fastholde indekset i datasættet.

En reliabilitetsanalyse af indekset "Skolebibliotekarernes tilgængelighed" har følgende Cronbach's alpha:

Tabel 32: Cronbach's alpha for indekset "Skolebibliotekarernes tilgængelighed"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,704	3

Note 32: N=7835

Med en Cronbach's alpha på 0,704 er tilliden til indekset ganske fin, og vi har derfor valgt at fastholde indekset "Skolebibliotekarernes tilgængelighed". Indekset er konstrueret således, at jo højere værdi på indekset, desto mere tilgængelige opleves skolebibliotekarerne.

Vi har også foretaget en faktoranalyse af spørgsmålsbatteri 14c, hvor vi har spurgt til, hvad eleverne synes om bibliotekarerne på biblioteket uden for skolen. Igen her skal vi være opmærksomme på, at det kun er elever, der har sagt, at de tager på biblioteket uden for skolen, som har besvaret spørgsmålene. Der er i alt 8721-5726=2995 elever som har sagt, at de ikke besøger biblioteket uden for skolen.

Tabel 33: Faktoranalyse af spørgsmål i spørgsmålsbatteri 14c.

	Rotated Component Matrix^a	
	Component 1	Component 2
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - Jeg kan spørge dem om alting	0,792	-0,160
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - De giver mig gode tips til bøger	0,803	-0,187
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - De lytter	0,860	-0,170
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - De er venlige	0,782	-0,086
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - De er ikke så nemme at snakke med	-0,191	0,672
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - Jeg snakker aldrig med bibliotekarerne	-0,115	0,847
14c. Hvad synes du om bibliotekarerne på biblioteket uden for skolen? - Jeg ser dem næsten aldrig	-0,119	0,814

Note 33: Varimax rotation. N=5726

Vi kan se i Tabel 33, at faktoranalysen siger, at der skal være to forskellige dimensioner. Vi har valgt at døbe den første dimension for "folkebibliotekarernes imødekommenhed" og den anden dimension er døbt "folkebibliotekarernes tilgængelighed".

I en reliabilitetsanalyse af dimensionen "folkebibliotekarernes imødekommenhed" er Cronbach's alpha:

Tabel 34: Cronbach's alpha for indekset "folkebibliotekarernes imødekommenhed"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,839	4

Note 34: N=5726

For indekset "folkebibliotekarernes imødekommenhed" er Cronbach's alpha 0,839, hvilket betragtes som en høj troværdighed, og dermed vil vi beholde indekset i datasættet.

I en reliabilitetsanalyse af dimensionen "folkebibliotekarernes tilgængelighed" er Cronbach's alpha følgende.

Tabel 35: Cronbach's alpha for indekset "folkebibliotekarernes tilgængelighed"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,706	3

Note 35: N=5726

For indekset "folkebibliotekarernes tilgængelighed" er Cronbach's alpha 0,706, hvilket betragtes som acceptabelt, og vi vil dermed beholde indekset i datasættet.

I en faktoranalyse af spørgsmålsbatteri 15, hvor vi har spurgt børnene om, hvad de synes om de historier og romaner, de læser i dansk, har vi fået følgende resultat:

Tabel 36: Faktoranalyse af spørgsmål i spørgsmålsbatteri 15

	Rotated Component Matrix^a		
	1	2	3
15. Hvad synes du om de historier og romaner, du læser i dansk? (Sæt et X i hver linje) - Jeg synes, de er gode/spændende	0,832	-0,106	-0,032
15. Hvad synes du om de historier og romaner, du læser i dansk? (Sæt et X i hver linje) - Jeg synes, de lærer mig noget om andre menneskers tanker og følelser	0,762	0,067	0,079
15. Hvad synes du om de historier og romaner, du læser i dansk? (Sæt et X i hver linje) - Jeg synes, de er svære	0,030	-0,028	0,981
15. Hvad synes du om de historier og romaner, du læser i dansk? (Sæt et X i hver linje) - Jeg får lyst til at læse	0,748	-0,119	0,091
15. Hvad synes du om de historier og romaner, du læser i dansk? (Sæt et X i hver linje) - Jeg synes ikke, man kan bruge dem til noget	-0,644	0,343	0,221
15. Hvad synes du om de historier og romaner, du læser i dansk? (Sæt et X i hver linje) - Jeg kan bedre lide, når jeg selv bestemmer, hvad jeg skal læse	-0,102	0,965	-0,042

Note 36: Varimax rotation. N=8721.

I Tabel 36 kan vi se, at der er 3 dimensioner i spørgsmålsbatteri 15. Den anden og tredje dimension består af hver et item, mens den første dimension udgøres af 4 items. Vi har valgt at døbe den første dimension for "brugbarhed". Den anden dimension er døbt "selvstændighed", mens den sidste dimension er døbt "sværhed".

En reliabilitetsanalyse kræver flere items i dimensionen og da der kun er et item i dimensionerne "selvstændighed" og "sværhed", kan vi kun lave en reliabilitetsanalyse på dimensionen "brugbarhed". Cronbach's alpha er brugt som mål for reliabiliteten på indekset.

Tabel 37: Cronbach's alpha for indekset "Brugbarhed"

Reliability Statistics

Cronbach's Alpha	N of Items
0,753	4

Note 37: N=8721

Vi kan se i Tabel 37 at indekset "brugbarhed" har en acceptabel Cronbach's alpha, og vi har derfor fastholdt indekset i datasættet.

I en faktoranalyse af spørgsmålsbatteri 16, hvor vi har spurgt børnene om, hvor meget tid de bruger på følgende ting uden for skolen på en normal hverdag, har vi fået følgende faktorer.

Tabel 38: Faktoranalyse af spørgsmål i spørgsmålsbatteri 16

Pattern Matrix^a

	Component		
	1	2	3
16. Hvor meget tid bruger du cirka på følgende ting uden for skolen på en normal hverdag? - Jeg ser TV	0,719	-0,199	0,144
16. Hvor meget tid bruger du cirka på følgende ting uden for skolen på en normal hverdag? - Jeg ser film, video eller programmer på (fx Netflix, YouTube, HBO, ViaPlay)	0,527	0,430	-0,069
16. Hvor meget tid bruger du cirka på følgende ting uden for skolen på en normal hverdag? - Jeg spiller spil på computer, telefon, tablet, konsol (fx Fifa, Sims, Minecraft, GTA, Hay Day)	0,836	-0,080	-0,133
16. Hvor meget tid bruger du cirka på følgende ting uden for skolen på en normal hverdag? - Jeg læser historier og tekster på Internettet (fx netavis, blogs, e-bøger, YouTube)	0,198	0,319	0,576
16. Hvor meget tid bruger du cirka på følgende ting uden for skolen på en normal hverdag? - Jeg læser historier og tekster i bøger og blade	-0,097	-0,127	0,919
16. Hvor meget tid bruger du cirka på følgende ting uden for skolen på en normal hverdag? - Jeg bruger sociale medier (fx Facebook, Instagram, Twitter, SnapChat)	-0,204	0,976	-0,020

Note 38: Promax rotation. N=8721

I Tabel 38 kan vi se, at faktoranalysen har givet os 3 dimensioner, men de er ikke klare. Det skyldes, at spørgsmål 2 indgår både i dimension 1 og 2, og dermed påvirkes besvarelsen af spørgsmålet af to dimensioner. Vi har valgt, at lade spørgsmål 2 indgå i begge indeks ligeværdigt med de andre spørgsmål i dimensionen. Vi har ikke rigtigt kunne navngive den første eller anden dimension, så har vi bare døbt dem faktor_16_1 og faktor_16_2. Den sidste dimension har vi døbt "prioriteret læsning".

I en reliabilitetstest for indekset "faktor_16_1" får vi følgende Cronbach's alpha:

Tabel 39: Cronbach's alpha for indekset "Faktor_16_1"

Reliability Statistics

Cronbach's Alpha	N of Items
0,541	3

Note 39: N=8721

I Tabel 39 kan vi se, at indekset "Faktor_16_1" har en Cronbach's alpha på 0,541, hvilket er i den lave ende af, hvad der er acceptabelt. Vi har dog valgt at beholde alle indeks.

I en reliabilitetsanalyse for indekset "faktor_16_2" får vi følgende Cronbach's alpha:

Tabel 40: Cronbach's alpha for indekset "Faktor_16_2"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,465	2

Note 40: N=8721

I Tabel 40 kan vi se, at indekset "Faktor_16_2" har en Cronbach's alpha på 0,465, hvilket er for lavt til at vi kan acceptere det. Spørgsmålene kredser som de sociale medier og film, hvilket ikke nødvendigvis er helt det samme. Spørgsmålene drejer sig om de sociale medier; fællesnævneren mellem dem kunne være, at det er netbaserede medier. Men når vi ser på reliabiliteten på indekset, så udelukker de spørgsmål gensidigt hinanden, da man reelt set kun kan bruge det ene type internet brug ad gangen. Dermed bliver tilliden til indekset lavt. Vi har dog stadigvæk indekset i vores datasæt, da det kan give en indikator for internetbrugen hos børnene.

Tabel 41: Cronbach's alpha for indekset "prioriteret læsning"

Reliability Statistics	
Cronbach's Alpha	N of Items
0,405	2

Note 41: N=8721

I Tabel 41 har indekset "prioriteret læsning" en Cronbach's alpha på 0,405, hvilket er i den lave del. Igen er forklaringen, at de to spørgsmål måler det samme, men er gensidigt udelukkende, da man reelt set kun kan læse en tekst af gangen. Vi har valgt at fastholde indekset, da det kan give en ide om, hvor meget tid børnene bruger på læsning generelt. Ligeegyldigt om det er i digitalt eller analogt format.

Indekskonstruktion

Til sidst vil vi gennemgå hvordan indeksene er konstrueret, og hvad fordelene er ved denne metode.

Vi har i undersøgelsen om børns læsevaner i fritiden valgt at konstruere indeks for de dimensioner vi fandt i faktoranalyserne af spørgsmålsbatterierne. Vi har valgt dette, da det skaber mere variation i vores data end de oprindelige spørgsmål, samt vi vil kunne lave mere interessante analyser. Indeksene er konstrueret således at indeksværdien er gennemsnitsværdien af svarværdierne af alle spørgsmål i indekset. Dette giver den fordel at man kan relatere indeksværdien til svarkategorierne til spørgsmålene. Det vil sige at hvis man har scoret 4 i indeksværdi på et indeks, hvor spørgsmålenes svarkategorier var "Meget enig"=4, og "Meget uenig"=1, så betyder det at vi via indeksværdien kan se at respondenteren er meget enig. Det betyder

ligeledes at respondenterne har svaret "Meget enig" på alle spørgsmål. Lignende gør sig gældende længere nedad indekset. Det kræver implicit at afstanden imellem hver svarkategori er lige stor. Variationen af indeksværdierne på et indeks med 3 spørgsmål, hvor hvert spørgsmål har 4 svarkategorier giver 10 kombinationsmuligheder.

En anden variant for konstruktion af indeks er ved at bruge faktorscorer. Denne metode antager at afstanden imellem alle svarkategorierne i hvert spørgsmål ikke er lige stor. Dette betyder at variationen afhænger af hvad man har svaret til de enkelte spørgsmål. Dermed vil vores indeks med hver 4 svarkategorier have 64 kombinationsmuligheder. Ergo vil faktorscorerne i højere grad ligne en kontinuert skala. Faktorscorerne har den udfordring, at denne score ikke er relaterbar til svarkategoriscoren, da indekset vil blive standardiseret omkring 0. Dermed kan en respondent godt have en negativ score på indekset, og det vil gøre fortolkningen meget svær og dermed ryger intuitionen væk.

Vi har derfor valgt at bruge gennemsnitsbetragtningerne, da gevinsten ved at kunne relatere indeksværdierne til svarkategorierne er meget større end den ekstra præcision, man får ved at bruge faktorscorerne. Med andre ord har vi valgt intuitionen frem for den matematiske præcision. Vi ønsker at vise, hvordan den relative holdning er blandt vores respondenter. Det vil sige, at vi med gennemsnit som indeks vil få en ide om, hvem der score højest på et indeks, og hvem der ikke gør. Der er en kraftig positiv sammenhæng imellem faktorscorerne og gennemsnitsværdierne.

Bibliografi

Bundsgaard, J., & Puck, M. R. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. Odense: DPU, Aarhus Universitet & Center for Anvendt Skoleforskning ved University College Lillebælt.

Rusmann, A., Ahlmann Plesner, M., & Puck, M. R. (2017). *Rapporten om indsamling og behandling af spørgeskemadata i Læremiddelundersøgelsen i danskfaget*. København: DPU, Aarhus Universitet.