

# PISA 2022

# Data og metode

Delrapport



*PISA 2022 Data og metode – Delrapport*

© VIVE og forfatteren, 2023

e-ISBN: 978-87-7582-277-5

Forsidegrafik: Hanne Bang Christensen/VIVE

Projekt: 301336

Finansiering: Børne- og Undervisningsministeriet

## **VIVE**

Det Nationale Forsknings- og Analysecenter for Velfærd

Herluf Trolles Gade 11

1052 København K

[www.vive.dk](http://www.vive.dk)

VIVEs publikationer kan frit citeres med tydelig kildeangivelse.



## Forord

OECD-programmet PISA (Programme for International Student Assessment) er et projekt, der har til formål at undersøge, hvor godt 15-årige elever er forberedt på at møde udfordringerne i dagens videns- og informationssamfund.

PISA 2022 gennemføres i Danmark af et konsortium bestående af VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd (VIVE) og Danmarks Statistik (DST). Fageksperter fra VIA University College, UCL Erhvervsakademi og Professionshøjskole samt Københavns Professionshøjskole er tilknyttet undersøgelsen. Konsortiet har nedsat en bestyrelse bestående af seniorforsker Vibeke Tornhøj Christensen, National Project Manager for PISA (VIVE), seniorforsker Louise Beuchert (VIVE), specialkonsulent Monika Klingsbjerg-Besrechel, datamanager for PISA (DST), forskningschef Andreas Rasch-Christensen (VIA), forskningschef Thomas Illum Hansen (UCL) og forskningschef Erik Caparros Højbjerg (KP).

Børne- og Undervisningsministeriet finansierer PISA-undersøgelsens gennemførelse, og en repræsentant fra Styrelsen for Undervisning og Kvalitet er medlem af PISA Governing Board (PGB), hvor deltagerlandene fastlægger de overordnede rammer for undersøgelsen sammen med OECD. Styrelsen for Undervisning og Kvalitet deltager desuden i konsortiebestyrelsesmøderne og bidrager til kvalitetssikringen af undersøgelsen i Danmark.

Undersøgelsens design og gennemførelse forestås af et internationalt PISA-konsortium bestående af internationale kontraktholdere udvalgt af OECD. Det internationale PISA-konsortium har trukket på internationale ekspertgrupper og faglige referencegrupper.

Ud over forskerne har personale og ikke mindst 6.200 elever ved 347 uddannelsesinstitutioner, repræsentativt udvalgt i Danmark, medvirket aktivt i undersøgelsen, der ikke havde været mulig uden dem. De takkes for deres bidrag.

*Hans Hummelgaard*

Forsknings- og analysechef for VIVE Effektmåling



# Indholdsfortegnelse

---

1	Metode og datakvalitet i PISA 2022	5
1.1	PISA-målgruppen	5
1.2	Testdesign, testopgaver og spørgeskemaer	5
1.3	Stikprøveudtrækket i PISA	8
1.4	Deltagelse og populationsdækning	10
1.5	Den praktiske gennemførelse af dataindsamlingen	15
1.6	Pilotundersøgelse	15
1.7	Datakvalitet	16
<hr/>		
	Litteratur	20

# 1 Metode og datakvalitet i PISA 2022

Designet i PISA 2022 er, som i de tidligere PISA-runder (2018, 2015, 2012, 2009, 2006, 2003 og 2000), fastlagt af det internationale konsortium i samarbejde med repræsentanter og eksperter fra de enkelte lande. Data skal være sammenlignelige mellem alle deltagerlandene, og samtidig skal det enkelte land kunne sammenligne resultaterne over tid. Designet i PISA 2022 er derfor i høj grad magen til de tidligere runder. Målgruppen er igen i 2022 elever, der på testafholdelsestidspunktet er 15-16 år gamle. Eleverne er alle fordelt på et repræsentativt udsnit af de danske uddannelsesinstitutioner og dækker friskoler og private grundskoler, folkeskoler, efterskoler, specialskoler, gymnasier og erhvervsskoler.

I det følgende vil de praktiske og tekniske forhold omkring PISA-testen i Danmark beskrives, herunder hvordan eleverne udvælges, og hvordan testen gennemføres på skolerne. For mere uddybende tekniske informationer angående testopgaver, stikprøve, bortfald og vægtning henvises til PISA 2022 Technical Report (OECD, under udgivelse).

## 1.1 PISA-målgruppen

Målgruppen i PISA 2022 er alle uddannelsessøgende personer født i år 2006. Det vil sige unge, som på testafholdelsestidspunktet er 15-16 år gamle og under uddannelse. I rapporten benævnes alle elever som 15-årige. I Danmark deltog ca. 7.800 unge fordelt på 347 uddannelsesinstitutioner, som omfatter både grundskoler og ungdomsuddannelser. 6.200 elever er testet i matematik, læsning og naturfag, og det er denne stikprøve, der behandles i denne rapport. De deltagende elever gik på klassetrin fra 7. klasse og opefter. Fordelingen af elever på de forskellige klassetrin kan ses i Tabel 1.1.

**Tabel 1.1** Fordelingen af elever på klassetrin

Klassetrin	Elever (procent)
7. klasse	0,2
8. klasse	9,7
9. klasse	89,7
10. klasse	0,4
Ungdomsuddannelser	0,1
I alt	100,0

Kilde: Westat (2023).

## 1.2 Testdesign, testopgaver og spørgeskemaer

PISA-undersøgelsen består for eleverne af en test i matematik, læsning og naturfag samt et spørgeskema om baggrundsforhold og holdninger. Herudover skal skoleledelsen fra hver enkelt skole

udfylde et spørgeskema om forhold, der vedrører skolens økonomi og personale. For mere information om udviklingen af spørgeskemaer, se PISA 2022 Assessment and Analytical Framework (OECD, 2023).

## 1.2.1 Testdesign og testopgaver

Datagrundlaget for PISA 2022 er en computerbaseret test, som består af en række praktiske opgaver inden for fagdomænerne matematik, læsning og naturvidenskab. Hver gang PISA gennemføres, gentages der en test i hver af disse domæner, men hovedfokus roterer mellem domænerne, hver gang PISA gennemføres. I 2022 er matematik hoveddomæne og har derfor størst fokus. Testen i 2022 er sammensat af:

- Matematik (hoveddomæne)
- Læsning og naturfag (underdomæner)
- Kreativ tænkning (innovativt domæne)
- Finansielle kompetencer (international option, med ekstrastikprøve<sup>1</sup>, som kan tilvælges).

Hver elev skal gennemføre 2 timers PISA-test efterfulgt af et baggrundsspørgeskema. Testen er sammensat af opgavesamlinger på hver 30 minutter fra de forskellige domæner, hvor kombinationen af opgaver varierer fra elev til elev. PISA-testen er udviklet til at vurdere et lands samlede præstation og ikke til at vurdere den enkelte elev. Således gennemfører hver elev ikke præcis de samme opgaver, og designet i PISA tillader derfor en meget bedre dækning af indholdet med et større materiale.

### Adaptive testning siden PISA 2018

Fra PISA 2018 blev der indført adaptive test for hoveddomænet læsning. Designet er trinvist (multistage adaptive testing) og bygger på et design, der med succes har været anvendt i OECD Programme for International Assessment of Adult Competencies (PIAAC). Multistage adaptive testing (MAT) betyder, at der på forskellige trin, på baggrund af besvarelse af en samling af opgaver, tildeles en ny samling opgaver med enten høj eller lav sværhedsgrad. I hver samling opgaver er der både test med åbne og lukkede besvarelser. Den adaptive vurdering sker på baggrund af de opgaver med lukkede besvarelser, der ikke skal vurderes manuelt. Adaptive test medfører en meget højere målepræcision, især for de højest og lavest præsterende elever. Dette design er videreført i PISA 2022 og udvidet til også at omfatte matematik, der er hoveddomæne i PISA 2022. For yderligere information om designet og den adaptive test, se PISA 2022 Technical Report (OECD, under udgivelse) samt PISA 2022 Integrated design (OECD, 2022).

---

<sup>1</sup> I alt deltog ca. 3.600 elever i finansielle kompetencer. Heraf deltog 2.030 elever også i hoveddomænet og underdomæner, mens et ekstrasample på ca. 1.570 elever udelukkende deltog i finansielle kompetencer. Disse 1.570 elever indgår ikke i data-materialet i hovedrapporten.

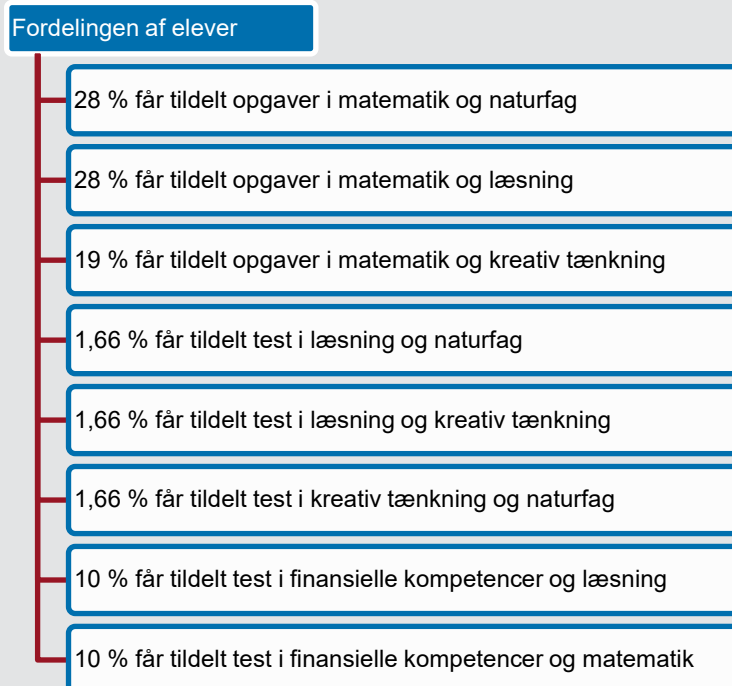
**Tabel 1.2 Oversigt over domæner og antal opgavesamlinger i PISA 2022**

Opgavesamlinger i PISA 2022		
Domæner	Pilot	Hovedundersøgelsen
Matematik NYE opgaver	12 opgavesamlinger a 30 minutter	I alt 60 minutters adaptiv test baseret på 15 opgavesamlinger a 30 minutter
Matematik TREND-opgaver (opgaver fra tidligere PISA-runder, der gentages i PISA 2022)	7 opgavesamlinger a 30 minutter	
Læsning	I alt 60 minutters adaptiv test baseret på 12 opgavesamlinger a 30 minutter (identisk med materialet fra 2018, men reduceret med ca. 25 pct.)	
Naturvidenskab	6 opgavesamlinger a 30 minutter	
Kreativ tænkning NYE opgaver	5 opgavesamlinger a 30 minutter	4 opgavesamlinger a 30 minutter
Finansielle kompetencer	2 opgavesamlinger a 30 minutter	2 opgavesamlinger a 30 minutter

Kilde: OECD (2022).

En oversigt over domæner og antal opgavesamlinger kan ses i Tabel 1.2. De forskellige opgavesamlinger fra de fem domæner samles i en række testforms, som administreres til en fastsat andel af de udtrukne elever. En oversigt over fordelingen kan ses i Figur 1.1.

**Figur 1.1 Fordeling af opgaver og domæner**



Kilde: OECD (2022).

### **Gentagelse af gamle testopgaver sikrer sammenligninger over tid**

Som det ses i Tabel 1.2, er testen – som i de tidligere runder – sammensat af både nye og ”gamle” (trend) opgaver samt forskellige kombinationer af de tre domæner (se Figur 1.1). Gentagelse af de samme opgaver over tid muliggør sammenligninger af resultater fra tidligere runder. Fra PISA 2015 skete der en ændring i designet og sammensætningen af opgaverne, som fremadrettet har gjort sammenligningen over tid stærkere, da bl.a. antallet af opgaver, der gentages, bliver øget. For uddybende informationer om skalering, se PISA 2022 Technical Report (OECD, under udgivelse).

Sammenligningen over tid ved hjælp af trendopgaverne muliggør en estimation af den absolutte udvikling i elevernes kompetencer. Ved at se på, hvordan elever i forskellige PISA-runder har klarer sig i de samme testopgaver, kan man udregne den absolutte udvikling i elevernes testscorer. Men da alle kompetencer kan ændre betydning over tid, vil der altid være en usikkerhed ved den absolutte ændring, jo længere tilbage man analyserer tidsrækken.

### **Tilføjelse af nye testopgaver sikrer relevans**

Trendopgaverne har også en anden vigtig funktion, da de sammen med de nye spørgsmål indgår i den score, der dannes *relativt* for hvert land. Viden og krav til kompetencer ændres over tid. De nye opgaver afspejler de kompetencer og den viden, der menes at være mest relevant i dag. Når et land flytter sig *relativt*, altså forholdsvist i sammenligning med andre lande, er det udtryk for, at landet er mere eller mindre parat til at møde de krav, der stilles i dag – og altså er bedre eller dårligere til at løse de opgaver, der måler nutidens kompetencer.

## **1.2.2 Spørgeskemaer**

Ud over selve testopgaverne indgår der i datagrundlaget for PISA 2022 en lang række baggrundsoplysninger om den enkelte elev og den skole, eleven går på. Disse oplysninger er hentet via elektroniske spørgeskemaer til elever og til skolelederne på de deltagende skoler.

## **1.3 Stikprøveudtrækket i PISA**

Grundprincippet i stikprøveudtagningen er, at PISA-stikprøver udtrækkes i to trin. Først udtrækkes et antal skoler blandt alle landets skoler, der potentielt kunne tænkes at have indskrevet elever født i 2006. Skolerne udtrækkes med en sandsynlighed, der er proportional med antallet af elever i målgruppen, der går på skolen. Hvis én skole har dobbelt så mange elever i målgruppen som en anden, har denne skole således dobbelt så stor sandsynlighed for at blive udtrukket som den anden. Et fastlagt antal elever fra hver enkelt skole udvælges efterfølgende tilfældigt. Jo færre elever, der er på en skole, jo større sandsynlighed vil den enkelte elev have for at blive udtrukket, da der ikke er så mange at vælge imellem. Når de små skoler har mindre sandsynlighed for at blive udtrukket end de store, mens eleverne på de små skoler til gengæld har en større sandsynlighed for at blive udvalgt end de enkelte elever på de store skoler, bliver det samlede udtræk selvvejet og repræsentativt.

### **Oversampling af skoler med mange elever med anden etnisk baggrund end dansk**

Danmark har i PISA 2022 – som i PISA 2009, PISA 2012, PISA 2015 og PISA 2018 – valgt at have særlig fokus på elever med anden etnisk baggrund end dansk. For at have et så stærkt datagrundlag som muligt er der udvalgt flere skoler med mange elever med anden etnisk baggrund



end dansk, og på et udvalg af de deltagende skoler er der udtrukket flere elever med anden etnisk baggrund end dansk. Det gøres ved at lave følgende stratificering af skoler før udtrækning: Alle skoler i populationen er delt op i fire strata, som er defineret ved en kombination af antal og andel elever på skolen, som har en anden etnisk baggrund end dansk:

1. **Høj:** Skoler med mindst fire elever med anden etnisk baggrund end dansk, og hvor andelen af elever med anden etnisk baggrund end dansk udgør mindst 33 %
2. **Mellem:** Skoler med mindst fire elever med anden etnisk baggrund end dansk, og hvor andelen af elever med anden etnisk baggrund end dansk udgør mindst 10 %, men færre end 33 %
3. **Lav:** Skoler, hvor antallet af elever med anden etnisk baggrund end dansk var mellem 1 og 3, og/eller hvor andelen af elever med anden etnisk baggrund end dansk udgjorde mere end 0, men under 10 %
4. **Ingen:** Skoler uden elever med anden etnisk baggrund end dansk.

I nogle strata trækkes en større andel af skolerne end i andre strata. Derfor foretages vægtningen også stratum for stratum, så hvis der i et stratum er valgt en større andel af skolerne, vægter de tilsvarende mindre i tabellerne.

**Tabel 1.3 Fordeling af skoler og elever – i populationen og i stikprøven**

Stratum	Population		Stikprøve (estimeret antal baseret på institutionslisten fra Danmarks Statistik)	
	Antal elever	Antal skoler	Antal elever	Antal skoler
01 = Høj	3.697	108	1.975	98
02 = Mellem	19.443	331	2.363	88
03 = Lav	29.595	700	2.977	125
04 = Ingen	11.133	1.101	1.123	100
I alt	63.869	2.240	8.438	411

Kilde: Westat (2023).

For hver skole i stratum 01 udtrækkes tilfældigt 36 elever. For hver skole i stratum 02 udtrækkes tilfældigt 24 elever blandt de danske elever, og 18 elever med anden etnisk baggrund end dansk tilføjes udtrækket. For hver skole i stratum 03 udtrækkes tilfældigt 30 elever blandt de danske elever, og 12 elever med anden etnisk baggrund end dansk tilføjes udtrækket. For hver skole i stratum 04 udtrækkes tilfældigt 34 elever blandt de danske elever, og 8 elever med anden etnisk baggrund end dansk tilføjes udtrækket.

På de skoler, hvor der er valgt flere elever med en anden etnisk baggrund end dansk, vægter disse elever tilsvarende mindre i tabeller og analyser, så alt er i balance.

Den efterfølgende vægtning tager således højde for det konkrete design med hensyn til, hvordan både skoler og elever er valgt i hvert stratum, og sikrer, at data er repræsentativ for populationen af 15-16-årige under uddannelse.

## Datagrundlag

Datagrundlaget for stikprøveberegningerne og udtrækket for både pilotundersøgelsen og hovedundersøgelsen er data fra Danmarks Statistik over alle uddannelsesinstitutioner og deres indskrevne elever. For hovedundersøgelsens vedkommende omfatter listen alle uddannelsesinstitutioner og antal indskrevne elever født i 2005 i Danmark pr. 30. september 2020, dvs. status efter begyndelsen på skoleåret 2020/2021. På basis af denne liste estimeres antallet af elever født i 2006 på de enkelte uddannelsesinstitutioner i skoleåret 2021/2022, hvor PISA fandt sted.

## 1.4 Deltagelse og populationsdækning

De tekniske standarder for PISA 2022 beskriver en række mål og betingelser i tilknytning til stikprøveetablering og deltagelsesprocent, der skal efterleves. Tabel 1.4 viser minimumskravene og de tilsvarende resultater for Danmark.

Et centralt element i PISA-stikprøvedesignet er, at der, som led i stikprøveudvælgelsen for hver af de udtrukne skoler, på forhånd udtrækkes en 1. reserveskole og en 2. reserveskole. Disse reserveskoler trækkes, så de udtrukne reserveskoler matcher en hovedskole ud fra samme principper og design som hovedstikprøvens. Disse bruges, hvis den oprindeligt udtrukne skole ikke deltager i undersøgelsen. Den negative betydning af eventuelt bortfald reduceres således væsentligt ved brug af erstatningsskoler, da de tre skoler er af samme type, ligger geografisk tæt på hinanden og har samme størrelse.

### I alt 347 skoler deltog i PISA 2022

Tabel 1.4 viser minimumskravene for deltagelse i PISA 2022. Af det oprindelige udtræk på 411 skoler var der 30 skoler, som viste sig ikke at have nogen elever i PISA-målgruppen. Det kan fx være skoler, som i princippet kunne have 15-årige elever, men som ikke havde det i det pågældende skoleår. 10 skoler blev fritaget, fordi de kun havde elever, der ville blive fritaget fra testen på grund af særlige undervisningsbehov. Af de resterende 371 skoler var der 46 skoler, som ikke ønskede at deltage. Det bragte antallet af deltagerskoler ned på 325. Oven i dette antal var der 22 reserveskoler, som deltog i stedet for de skoler, der ikke ønskede at deltage, hvilket bragte det samlede antal op på 347 deltagende skoler.

**Tabel 1.4 Minimumskrav for deltagelse i PISA 2022**

Vurderingsparameter	Minimumskrav	Status for Danmark 2022	Status for Danmark 2018	Status for Danmark 2015
Antal udtrukne skoler		411	396	400
Antal udtrukne skoler, der ikke længere havde PISA-elever		30	20	20
Antal udtrukne skoler, der blev fritaget		10	5	9
Antal udtrukne skoler, der deltager	150	325	328	327
Antal udtrukne skoler, der deltager i alt (inkl. reserveskoler)		347	344	331
Andel udtrukne skoler, der deltager (uvægtet)		87,6 %	88,41 %	88,14 %
Andel udtrukne skoler, der deltager (vægtet)	85 %	90,09 %	88,11 %	90,46 %

Vurderingsparameter	Minimums-krav	Status for Danmark 2022	Status for Danmark 2018	Status for Danmark 2015
Andel udtrukne skoler, der deltager, inkl. erstatningsskoler (uvægtet)		93,53 %	92,72 %	89,22 %
Andel udtrukne skoler, der deltager, inkl. erstatningsskoler (vægtet)	–*	96,20 %	93,34 %	92,03 %
Andel elever på udtrukne skoler, der deltager (uvægtet)		83,17	85,56 %	87,35 %
Andel elever på udtrukne skoler, der deltager (vægtet)	80 %	84,21	86,44 %	89,08 %
Populationsdækning (omvendt fritagelsesprocent)	95 % (5 %)	88,45 (11,55 %)	94,3 % (5,7 %)	94,96 % (5,04 %)

Note: \* Irrelevant, da kravet til opnåelsesprocenten uden erstatningsskoler er opfyldt.

Kilde: Westat (2023) og OECD (under udgivelse).

### Fritagelse af elever med særlige undervisningsbehov

Hvis elever har særlige undervisningsbehov (SEN, Special Education Needs) og efter individuel vurdering fra skolens personale, i samråd med Danmarks Statistik og testadministratoren, vurderes til ikke at ville kunne gennemføre PISA-testen, kan de blive fritaget fra deltagelse.

Hele skoler kan også blive fritaget, hvis skolen udelukkende har elever, der vil blive fritaget på grund af SEN, eller hvis skolen har et andet undervisningssprog end dansk.

Årsager, som kan begrunde, at elever bliver fritaget, er:

- Funktionelt handicap: Eleven har et moderat eller svært fysisk handicap.
- Kognitivt, adfærdsmæssigt eller følelsesmæssigt handicap: Bedømt på baggrund af vurdering fra kvalificeret personale har eleven et kognitivt, adfærdsmæssigt eller følelsesmæssigt handicap.
- Begrænset erfaring med testforløbets sprog: Eleven har ikke dansk som modersmål og har begrænsede færdigheder i dansk, herunder har modtaget undervisning i dansk i mindre end et år.
- Ord- eller talblind: Efter bedømmelse fra kvalificeret personale er eleven ord- eller talblind.

Tabel 1.5 viser fordelingen af fritagelse i Danmark. Her ses det, at elever, der af skolen vurderes til at være ord- eller talblinde i en sådan grad, at de skal fritages fra testen, udgør halvdelen af alle elever, der fritages.

Elever med særlige undervisningsbehov tilbydes en særlig en-times test kaldet UH (Une Heure). UH-forløbet består af en reduceret test samt et reduceret elevspørgeskema. Testen er konstrueret til at skulle vare en time og består af et udvalg af de lettere tilgængelige opgaver fra den ordinære PISA-test. Elevernes besvarelser og bedømmelsen heraf indgår på lige fod med de øvrige elevers. UH-forløbet gennemføres på elevernes præmisser, og der gøres brug af eksempelvis ekstra pauser, fysisk aktivitet mv. UH-forløbet benyttes både på specialskoler og almindelige skoler.

**Tabel 1.5 Fritagelse i Danmark**

Vurderingsparameter	2022	2022	2018	2018	2015	2015
	Antal fritaget	Procent ud af alle, der er fritaget	Antal fritaget	Procent ud af alle, der er fritaget	Antal fritaget	Procent ud af alle, der er fritaget
Funktionelt handicap	14	1,55	15	3,38	18	3,42
Kognitivt, adfærdsmæssigt eller følelsesmæssigt handicap	330	36,59	179	40,32	281	53,42
Begrænset erfaring med testforløbets sprog	102	11,31	88	19,82	156	29,66
Ord- eller talblind	456	50,55	162	36,49	71	13,50
Total	902	100	444	100	526	100

Ud over implementeringen af UH-testen er der et øget fokus på at etablere en god og tæt dialog med de enkelte skoler med information og rådgivning om, hvilken type elever der kan fritages fra PISA-testen, og hvorvidt de kan deltage i UH-testen. På denne måde bliver uklarheder håndteret tidligt i forløbet for at sikre, at kun elever, der reelt ikke *kan* deltage i PISA på grund af SEN, bliver fritaget.

### Betydning af stigning i fritagne elever

I alt er 11,55 % af eleverne fritaget fra testen, hvilket er mere end de tekniske standarder tillader (5 %) (OECD 2023, Vol. I, Reader's Guide). Overskridelsen beror på især en øgning i antallet af ordblinde, der fritages fra deltagelse i PISA (Tabel 1.5), da de ikke kan benytte de hjælpemidler, de ellers benytter i skolehverdagen. Der er næsten sket en tredobling af antallet af elever, der fritages pga. ordblindhed: fra 162 elever i 2018 til 456 elever i 2022.

OECD (2023, Reader's Guide og Annex A2) diskuterer betydningen af dette og vurderer, at en fritagelsesprocent på dette niveau kan overestimere et lands gennemsnit i hhv. matematik, naturfag og læsning med op mod 11 point.<sup>2</sup> Beregningen antager en moderat høj korrelation mellem elevens faglige niveau og det at blive fritaget fra test, og at de elever, der ikke deltager i PISA-testen, ville have opnået lavere resultater, end de elever, der deltager.

Det er dog vores vurdering, at betydningen af den øgede fritagelse ikke flytter det samlede resultat mere end op til 6 point, da fritagelsen af danske elever primært skyldes, at eleverne ikke kan få de hjælpemidler stillet til rådighed til gennemførelse af testen, som de er vant til, og ikke, at de er så fagligt svage, som OECD lægger til grund for deres beregninger (se analyse i næste afsnit).

Havde fritagelsesprocenten i Danmark i PISA 2022 været 5,7 % som i PISA 2018, ville de gennemsnitlige resultater, baseret på ovenstående antagelser og beregninger, være op til 6 point lavere i PISA 2022 end de afrapporterede domænegennemsnit i PISA 2022-rapporten. For at sætte disse tal i kontekst kan man se på, at hvert kompetenceniveau har et spænd på ca. 62 point i matematik, ca. 73 point i læsning og ca. 75 point i naturfag (OECD 2023, Vol. I, Reader's Guide, Box 1). I sammenligning hermed er 6 point ikke ret meget.

<sup>2</sup> OECD har tidligere beregnet, at en fritagelsesprocent på 5 (10) % vil medføre en overestimering af et lands nationale gennemsnitscore på op mod 5 (10) point på PISA-skalaen. Se PISA 2015 Technical report (OECD, 2017) og PISA 2018 Results, Vol. I (OECD, 2019, s. 161). En fritagelsesprocent på 11 antages derfor at kunne medføre en overestimering på op til 11 point.

En anden fortolkning kan være i forhold til standardfejlen, altså usikkerheden, på de rapporterede PISA-resultater. Danske elevers gennemsnit i matematik har en standardfejl på 1,9 point og dermed er 95-procents-konfidensintervallet +/- 3,7 point omkring gennemsnittet på 489 point (Tabel 1.7). Den potentielle overestimering (at det danske resultat ville have været op til 6 point lavere, hvis der ikke havde været en stigning i fritagelsesprocenten) ligger således inden for usikkerhedsintervallet (da de to tals usikkerhedsintervaller ville overlappe).<sup>3</sup>

Man kan læse mere om, hvornår en pointændring vurderes som stor eller lille i OECD (2023, Vol. I, Reader's Guide, Box 1).

OECD (2023, Vol. I, Reader's Guide) skriver dog også følgende om bl.a. resultaterne for Danmark: "The Adjudication Group also noted that the bias associated with trend and cross-country comparisons might be smaller, if past data or data for other countries are biased in the same direction." Det formodes at være tilfældet for både danske data og data fra de lande, vi sammenligner med over tid (jf. at der også er en fritagelsesprocent over 5 blandt svenske og norske elever) (OECD 2023, Vol. I, Annex A2).

### **Analyse af de fritagne elever og vurdering af betydningen for PISA-resultaterne**

Det danske PISA-konsortium har undersøgt de fritagne elever for at vurdere betydningen af stigningen i fritagne elever og OECD's vurdering af den potentielle overestimering (se ovenfor).

Tabel 1.5 viser, at halvdelen af de fritagne elever i Danmark er ordblinde eller talblinde. Tabellen viser også, at næsten 3 gange flere elever er fritaget af denne årsag end i 2018.

Vi har analyseret data om de danske PISA elevers afgangskarakterer (Tabel 1.6) og sammenlignet følgende fire grupper:

- (1) alle elever, der deltog i PISA
- (2) elever, der deltog i PISA, som ikke er vurderet til at være ordblinde eller andre SEN-koder
- (3) elever, der deltog i PISA, og som er vurderet ord-/talblinde
- (4) elever, der er blevet fritaget fra PISA grundet ord-/talblindhed.

Tabel 1.6 viser, at stort set alle de fritagne elever med ord-/ talblindhed (440-441 elever ud af total 456 fritagne elever, jf. Tabel 1.5) gik til afgangseksamen i de bundne prøver i hhv. dansk og matematik. Tabellen viser også, at de elever, der er fritaget pga. ordblindhed eller talblindhed (række 4), har 9. klasse-afgangskarakterer i dansk og matematik på niveau med afgangskaraktererne fra de ord-/talblinde, der har gennemført PISA-testen (række 3). Det bekræfter, at de fritagne elever kan gennemføre en eksamen. Begge grupper har lavere gennemsnitlige afgangskarakterer (ca. 2 karakterpoint, afhængigt af prøveform) end de øvrige deltager elever i PISA. Det vurderes derfor, at de fritagne elever ville trække det nationale gennemsnit ned, hvis de ikke var blevet fritaget fra deltagelse i PISA-testen, men ikke i nær så høj grad, som OECD's worst case scenario.

Vores vurdering er derfor, at disse elever godt ville kunne gennemføre PISA-testen, men at de fritages, fordi PISA ikke stiller LST (læse-skriveteknologier) til rådighed til gennemførelse af testen. På den baggrund vurderer vi, at en overestimering på 11 point er for højt i en dansk sammenhæng.

---

<sup>3</sup> Et 95-procents-konfidensinterval på matematik [485-493] i PISA 2022. Justeres gennemsnittet med 6 point, da er 95-procents-konfidensintervallet [479-487].

**Tabel 1.6 9. klasse-afgangskarakterer opdelt på elever, der deltager eller er fritaget fra PISA 2022**

Elevgruppe	9. klasse-afgangseksamen i dansk		9. klasse-afgangseksamen i matematik	
	Antal	Gns.	Antal	Gns.
(1) PISA deltager	6.071	8,0	6.098	7,8
(2) PISA deltager uden en SEN-kode	5.599	8,2	5.611	8,0
(3) PISA deltager med en SEN-kode for ordblindhed	270	6,1	271	5,8
(4) PISA fritaget med en SEN-kode for ordblindhed	440	5,7	441	5,4

Anm.: Tabellen viser elevernes eksamensgennemsnit af de bundne prøver i hhv. dansk og matematik ved 9. klasse afgangseksamen. Hvis en elev mangler to eller flere prøvekarakterer, er vedkommendes eksamensgennemsnit ikke beregnet, og indgår derfor ikke i denne tabel (gælder for 20 elever). Derudover mangler der nogle elever, som det ikke er muligt at koble med de danske registre om afgangskarakterer. SEN-koden for ordblindhed omfatter også elever vurderet at have talblindhed.

Kilde: Egne beregninger.

Vi har også undersøgt hvad det potentielt betyder for de danske elevers PISA-gennemsnit i matematik, læsning og naturfag at udelade elever med ord-/talblindhed (se Tabel 1.7). Tabellen viser, for PISA matematikresultaterne, at hvis de ordblind elever, der har gennemført PISA-testen, udelades af det samlede resultat, da opnår Danmark et gennemsnit på 491 point, hvilket kun er 2 point højere end det afrapporterede på 489 point. Data om fritagelse viser, at der skulle have deltaget omkring 3 gange flere antal ordblind, end der deltog (Tabel 1.5). Vurderingen er derfor, at Danmarks nationale gennemsnit kan være overestimeret med omkring 6 point.

**Tabel 1.7 PISA-resultater opdelt på elevernes status for ordblindhed**

Elevgruppe	Pisa 2022	Matematik				Læsning				Naturfag			
		Antal elever	Gns.	Std. fejl	95 % konf.int.	Gns.	Std. fejl	95 % konf.int.	Gns.	Std. fejl	95 % konf.int.		
(1) Alle elever	6.200	489	1,9	485	493	489	2,6	484	494	494	2,5	489	499
(2) Elever, der ikke har SEN-kode for ordblindhed	5.926	491	1,9	487	495	492	2,4	487	497	497	2,3	492	502
(3) Elever, der har SEN-kode for ordblindhed (men alligevel deltager i PISA-testen)	274	448	4,6	439	457	430	6,2	418	442	432	7,9	417	447
Er gruppe (1) og (2) gns. signifikant forskellige?		Ja				Ja				Ja			

Anm.: Tabellen viser elevernes gennemsnit i hhv. matematik, læsning og naturfag i PISA 2022. SEN-koden for ordblindhed omfatter også elever vurderet at have talblindhed. Elevgruppe (3) indeholder elever, der har en SEN-kode for ord/talblindhed, men alligevel deltager i PISA-testen (fx ved at tage UH-testen), det er således muligt at beregne elevernes domænegennemsnit i PISA 2022 med og uden denne elevgruppe inkluderet.

Kilde: Egne beregninger i IDB Analyser. Id på elever med ordblindhed er leveret af DST.

## 1.5 Den praktiske gennemførelse af dataindsamlingen

Testperioden for PISA 2022 i Danmark var fra 7. marts til 1. maj 2022.

Testen, som finder sted på det enkelte uddannelsessted, foregår som ved afholdelse af en eksamen. Testafholdelsen gennemføres ikke af læreren, men under opsyn og vejledning af uddannede testadministratorer fra Danmarks Statistiks interviewerkorps. Testen og udfyldelse af baggrundsspørgeskemaet bliver afviklet på bærbare computere, som Danmarks Statistik stiller til rådighed. Data fra testen lagres på computerens harddisk for derefter at blive overført til den samlede database.

Danmarks Statistiks interviewere fungerer som testadministratorer på de udvalgte skoler. På hver skole er der udpeget en skolekontaktperson, som sørger for det praktiske omkring testen, herunder orientering af forældrene og skolen som helhed med hjælp fra foldere udleveret af Danmarks Statistik. Derudover har det internationale konsortium ansat to personer fra Danmark til at overvåge testproceduren på et antal tilfældigt udvalgte skoler. Disse testkvalitetsledere bliver oplært og aflønnet af det internationale konsortium, hvem de også rapporterer direkte til. Der er ikke rapporteret om problemer i Danmark, som betyder, at Danmark ikke lever op til de fastlagte standarder.

Selve PISA-testforløbet tager i alt ca. 3 timer og 45 minutter og forløber således:

**Tabel 1.8 PISA-testforløbet**

Aktivitet	Elevtid
Præsentation af materialerne	15 minutter (ca.)
Introduktion	5 minutter (ca.)
Første 60 minutter af testen	60 minutter (nøjagtig)
Kort pause	Generelt ikke mere end 5 minutter
Introduktion	5 minutter (ca.)
Sidste 60 minutter af testen	60 minutter (nøjagtig)
Pause	15 minutter
Elevspørgeskemaet	55 minutter (ca.)
Afslutning af testforløbet (indsamling og overførsel af data)	5 minutter (ca.)
I alt	3 timer og 45 minutter (ca.)

## 1.6 Pilotundersøgelse

Pilotundersøgelsen gennemføres for at afprøve procedurer forud for hovedundersøgelsen og for at indsamle viden om opgaverne, så opgaver, der ikke i tilstrækkelig grad opfanger faglig variation hos eleverne og kan sammenlignes på tværs af lande, kan sorteres fra. Desuden bliver procedurerne i de mange manualer afprøvet. Ydermere er målet med pilotundersøgelsen at forbedre grundlaget for den adaptive test. Pilotundersøgelsen var planlagt til at gennemføres i 2020, men måtte udskydes til 2021, pga. covid-19-relaterede skolenedlukninger i foråret 2020. Pilotundersøgelsen blev således gennemført i perioden 6. april til 31. maj 2021, hvor 37 skoler og 1.289 elever deltog. Målgruppen var her uddannelsessøgende født i år 2005.

## 1.7 Datakvalitet

Som beskrevet er der i alle undersøgelsens praktiske led etableret omfattende procedurer for at sikre tilfredsstillende data. Hvis kravene i de tekniske standarder er opfyldt, bliver data automatisk godkendt. Hvis nogle krav ikke er opfyldt, foretager det internationale konsortium og det pågældende land nærmere analyser af data, og en ekspertgruppe vurderer, om data kan godkendes eller ej, ligesom landene også kan afkræves yderligere dokumentation. Samlet set vurderes de danske data at være af høj kvalitet, og de er indgået i de internationale sammenligninger med et enkelt forbehold, da fritagelsesprocenten er højere, end de tekniske standarder foreskriver.

Pålidelighed, validitet og repræsentativitet er tre nøglebegreber i relation til datakvalitet. Hvorvidt data er repræsentative, kan man langt hen ad vejen opstille eksakte statistiske mål for, mens datas pålidelighed og validitet er en forklarings- og dokumentationssag.

Hvis datapålideligheden skal være god, må der ikke være opstået fejl, som betyder, at de indsamlede data giver en dårlig beskrivelse af virkeligheden. Høj datapålidelighed er en forudsætning for, at data kan bruges til at drage holdbare konklusioner.

### 1.7.1 Testgennemførelsen

Pålideligheden i datagennemførelsen er sikret ved, at der bl.a. er udarbejdet adskillige vejledninger og manualer, som sikrer en ensartet og korrekt procedure i alle lande og på alle skoler, uafhængigt af, hvilken testadministrator der har forestået testen.

Der er bl.a. udarbejdet følgende vejledninger og manualer:

- Testmanual
- Vejledning til skolekontaktperson
- Manual til stikprøveudtrækning og datamanagement.

Alle personer, der har arbejdet med dataoparbejdningen (data- og sampling-manager), har modtaget undervisning fra det internationale konsortium, og de 80 testadministratorer, der varetog selve testafholdelsen på skolerne, deltog i et instruktionsmøde afholdt af datamanageren. På møderne blev hele testmanualen gennemgået, og forskellige situationer, der kunne opstå, blev diskuteret.

### 1.7.2 Kodning af de åbne besvarelser

En stor del af PISA-testen består af opgaver, som eleven skal besvare med egne ord. Bedømmelsen af elevernes åbne besvarelser kaldes kodning og foretages manuelt af mennesker. Al kodning foregår ud fra retningslinjer, der er fastlagt af det internationale konsortium.

18 kodemedarbejdere blev udvalgt ud fra deres faglige viden om de forskellige fagområder, som eleverne blev testet i, og deres kendskab til 15-årige elevers udtryksmåder. Kodemedarbejderne blev tilknyttet et specifikt fagdomæne (matematik, læsning, naturfag, finansielle kompetencer eller kreativ tænkning) og deltog i et intensivt kursus i kodning afholdt af en ekspert for hvert enkelt fagdomæne.



Alle kodemedarbejdere arbejdede under hele processen i grupper, hvor det skulle sikres, at de overholdt de internationale kodeprocedurer. Et stort antal elevbesvarelser på de åbne opgaver blev multikodet. Dette betyder, at kodemedarbejdernes resultater på samme elevbesvarelser blev sammenlignet med hinanden, og enkelte opgaver blev ligeledes sammenlignet med andre landes resultater for på denne måde at sikre kvaliteten og ensartetheden i kodningen på tværs af kodere og på tværs af lande. Hvis kodernes vurderinger ligger for langt fra hinanden, bliver de pågældende svar kodet på ny efter gentræning af koderne.

### 1.7.3 Testopgaverne

Kravet om pålidelighed (validitet) i undersøgelsens måleinstrumenter er udtryk for, om undersøgelsen faktisk afdækker det, den har til hensigt.

De testopgaver, der bruges i PISA, er udvalgt af ekspertpaneler i samarbejde med forskere i de enkelte lande, og der er brugt mange ressourcer på at sikre opgavernes validitet på tværs af tid og sted. De valgte opgaver må derfor siges at være det bedst mulige redskab til at måle elevernes kompetencer.

Når testopgaverne bruges i en dansk sammenhæng – ligesom i alle andre ikke-fransk/engelsktalende lande – skal opgaverne igennem en såkaldt dobbelt uafhængig oversættelsesproces med efterfølgende reconcilierung. Testopgaverne oversættes af to uafhængige oversættere: 1) fra engelsk til dansk og 2) fra fransk til dansk. De to oversættelser er derefter redigeret sammen til én "originalversion" af en person, som er kyndig i testkonstruktion. Sidstnævnte betegnes reconcilierung. Denne oversættelse bliver efterfølgende kontrolleret af nationale fageksperter for at sikre, at den valgte terminologi er i overensstemmelse med national praksis. Endelig bliver den reviderede originalversion kontrolleret af et oversættelsesfirma, der specialiserer sig i sproglig kvalitetskontrol, udvalgt af det internationale konsortium. Dette firma har gennem faglig ekspertise sammenholdt oversatte versioner fra flere lande. Eventuelle uoverensstemmelser på denne baggrund bliver diskuteret med de nationale eksperter, og den endelige version bliver om nødvendigt rettet til. Hvis eleverne ikke svarer på præcis den samme opgavetekst i de enkelte lande, undermineres sammenligningsgrundlaget og dermed validiteten af dette helt centrale element i undersøgelsen – derfor den meget grundige og omstændelige oversættelsesprocedure. Spørgeskema materialet gennemgår en lignende procedure med dobbelt uafhængig oversættelse og efterfølgende reconcilierung.

### 1.7.4 Repræsentativitet

Et yderligere krav til høj datakvalitet er, at den udtrykker præstationer og holdninger for hele den målgruppe, man ønsker at drage konklusioner om. Det betyder, at de 6.200 elever, der har deltaget i testen, skal repræsentere hele gruppen af uddannelsessøgende unge født i 2006. Tabel 1.9 viser en opgørelse af stikprøven fordelt på strata 01-04 (1: høj andel elever med anden etnisk baggrund end dansk, 2: mellem andel elever med anden etnisk baggrund end dansk, 3: lav andel elever med anden etnisk baggrund end dansk og 4: ingen elever med anden etnisk baggrund end dansk) sammenholdt med en tilsvarende fordeling for målgruppen. Vægtene tager højde for forskellige udtrækssandsynligheder og for bortfald på skole- og elevniveau.

**Tabel 1.9** Oversigt over strata og vægte i PISA 2022

Stratum	Elever, som har gennemført testen	Sum (elever i populationen)	Middelvægt for elever i stratum	Standardafvigelse af elevvægte i dette stratum
01 = Høj	484	977,67	2,02	0,66
01 = Høj (100 % ssh.) <sup>1</sup>	906	2.244,84	2,80	0,84
02 = Mellem	1.732	17.135,75	9,89	4,05
03 = Lav	2.193	26.296,27	11,99	5,09
04 = Ingen	885	10.291,28	11,30	6,62
I alt	6.200	56.945,82		

Note: <sup>1</sup>Totaludvalgt: I stratum 1 er nogle af skolerne så store, at de skal være med i stikprøven. De er derfor opregnet for sig.

Kilde: Westat (2023).

Når man beregner vægte til data for at korrigere for mange forskellige forhold (ulige udtræksandsynlighed på skole- og elevniveau samt ujævnt bortfald på både skole- og elevniveau), kan enkelte observationer (her: elever) få en meget høj vægt, hvis alle forhold for disse observationer trækker i samme retning. Dermed kan få observationer få stor indflydelse på det samlede resultat. Man vælger derfor ofte at lægge et loft over, hvor stor en vægt en enkelt observation kan få: såkaldt trimming. Dette gøres normalt ud fra en konkret analyse af data, og det gøres også på denne måde i PISA. Der var kun 5 elever i de danske data, som fik så stor vægt, at det blev vurderet nødvendigt at trimme vægten. Det viser, at man i Danmark – på trods af at der var flere hensyn at tage – er lykkedes med at designe et stikprøvedesign, der både er balanceret, robust og repræsentativt.

## 1.7.5 Testresultater og usikkerhed

PISA-resultaterne for hvert domæne opgøres i point på en skala, der er udviklet for hvert fagområde og designet til at vise de generelle kompetencer, der er testet i PISA. Gennemsnittet for hver af disse skalaer er sat til 500 point og med en standardafvigelse på +/- 100 point som gennemsnit for OECD-landene, første gang det pågældende domæne var hoveddomæne.

Ingen test er perfekt, og én test kan ikke måle alle færdigheder i brede fag som læsning, matematik og naturfag. I PISA, som for alle andre test og måleredskaber, kan der opstå målefejl og andre kilder til usikkerhed. Ovenfor er beskrevet, hvordan mulige usikkerheder er søgt reduceret, fx ved at bruge mange ressourcer på at sikre en pålidelig og systematisk gennemførelse af PISA-testen på samtlige deltagende skoler.

PISA-resultater er endvidere baseret på en stikprøve af de deltagende landes skoler og elever. I OECD's rapporter såvel som denne rapport er landenes resultater (fx gennemsnitlige matematikscore) derfor rapporteret sammen med et estimat for deres usikkerhed. Usikkerheden kan tænkes som spredningen på en stikprøves gennemsnit og kan beregnes og rapporteres som standardfejlen (standard error (s.e.)). Det rapporteres også, om beregnede forskelle (fx mellem to landes gennemsnitlige læsescore eller mellem drenge og piger) er statistisk signifikante. Derved redegøres for den statistiske usikkerhed på PISA-resultater, der er forbundet med, at resultaterne er baseret på en stikprøve. I Tabel 1.10 opsummeres de tre hovedkilder til usikkerhed i en stikprøvemåling som PISA, og hvordan disse behandles.

**Tabel 1.10 Usikkerheder i en stikprøvemåling, og hvordan disse behandles i PISA**

Usikkerhed	Beskrivelse og behandling
<p>Stikprøvefejl <i>Hvor stor spredning er der på stikprøven?</i></p>	<p>Stikprøvefejlen kan reduceres, jo større stikprøve, der udtrækkes. Jo større stikprøve, der udtrækkes, jo mere vil stikprøvegennemsnittene alt andet lige ligne hinanden (jo mindre spredning er der), og jo nærmere er stikprøvens gennemsnit på populationens gennemsnit. Spredningen på en stikprøves gennemsnit kan beregnes og kaldes standardfejlen (<i>standard error (s.e.)</i>).</p> <p>Metoderne bag stikprøveudtagningen i PISA sikrer, at stikprøven er repræsentativ for hele populationen og giver et validt estimat af gennemsnitsscoren og fordelingen, som kan generaliseres til hele populationen.</p> <p>Usikkerheden kan mindskes, hvis datagrundlaget øges (flere skoler og – dog i mindre grad – flere elever fra hver skole), men er minimeret så meget, som undersøgelsens budget og design tillader.</p> <p>For de fleste lande er standardfejlen på estimatet for et lands gennemsnitsscore 1-3 PISA-point. For OECD-gennemsnittet (som baserer sig på 37 stikprøver på landeniveau) er stikprøvefejlen reduceret til omkring 0,4 PISA-scorepoint. Se fx tabeller med landes gennemsnit og standardfejl i OECD (2023).</p>
<p>Målefejl <i>Hvad måler testopgaverne?</i></p>	<p>Der kan kun testes et begrænset antal af testopgaver, og det kan lede til spørgsmålet: Ville brugen af andre testopgaver have resulteret i andre resultater?</p> <p>Usikkerheden kan reduceres, jo flere opgaver, der tester samme færdighed. Usikkerheden er derfor større for den enkelte elevs testresultat (der kun ser fire bundter af testopgaver) end for landsgennemsnittene (som baserer sig på alle bundter med testopgaver). Usikkerheden mindskes også i takt med mængden af tilgængelig baggrundsinformation.</p>
<p>"Link"-fejl <i>Er en given score i PISA 2022 den samme som i PISA 2012?</i> <i>Relevant ved sammenligning over tid, hvor resultater "linkes" sammen med tidligere PISA-runder</i></p>	<p>Selvom alle PISA-undersøgelser bruger den samme skala og enhed (point), fastsat i den første runde (læsning blev fx defineret i PISA 2000), ændres testinstrumenterne og testspørgsmålene, stikprøverne samt nogle gange de statistiske modeller, der omregner elevbesvarelser til resultater på skalaen.</p> <p>For at gøre resultaterne direkte sammenlignelige over tid, er skalaerne nødt til at blive ækvivaleret. Dette betyder, at resultaterne bliver tilpasset, så de kan blive udtrykt efter den samme målestandard. "Link"-fejl kvantificerer den usikkerhed, der er forbundet med ækvivaleringen af skalaerne og dermed usikkerheden i trend mellem de forskellige målinger.</p> <p>Linkfejlen er usikkerheden omkring skalaværdierne og er derfor uafhængig af størrelsen på stikprøven af elever. For information om link-fejl, se PISA 2022 Technical Report (OECD) (Linking error, standard error and performance trend between PISA 2022 and previous PISA cycles).</p>

Kilde: OECD (2019) Vol. I, kapitel 2 og OECD (under udgivelse).

# Litteratur

Westat (2023). *PISA 2022, Weighting Summary Report, DENMARK*. Westat.

OECD (under udgivelse). *PISA 2022 Technical Report*. OECD Publishing.

OECD (2017). *PISA 2015 Technical Report*. OECD Publishing.

OECD (2019). *PISA 2018 Results, Vol. I: What Students Know and Can Do*. OECD Publishing.

OECD (2023). *PISA 2022 Integrated design*. OECD Publishing.

OECD (2023). *PISA 2022 Assessment and Analytical Framework*. OECD Publishing.



VIA University  
College



.....

Erhvervsakademi og  
Professionshøjskole



**DANMARKS  
STATISTIK**

**VIVÉ**