

National reference for de nationale test

Der er tre principielt forskellige repræsentative kvantitative opgørelse af læring i grundskolen i Danmark: De internationale undersøgelser (ILSA¹), Folkeskolens Prøver (FP) og De nationale test (DNT). De har alle som ambition at opgøre læring på nationalt niveau, men med hver deres mere specifikke formål, som også afspejler resultaternes tilgængelighed.

Særlige egenskaber ved DNT

De nationale test (DNT) har en række egenskaber i forhold til andre redskaber til opgørelser af faglig læring:

1. De anvendes på alle skoler for næsten alle elever i folkeskolen (i modsætning til ILSA),
2. De er baserede på måling ved brug af opgavebanker med lille udskiftning, dvs. det er de samme opgaver der anvendes fra år for år (modsat FP, hvor det er lutter nye opgaver fra år til år)
3. DNT-opgaver er af prøvet (i modsætning til FP) systematisk og kun de opgaver der passer til Raschmodellen er inkluderet i bankerne.
4. DNT kan (modsat FP og ILSA) benyttes flere gange pr. elev (både frivilligt og obligatorisk samt i dansk og matematik, obligatorisk på flere klassetrin) og da raschskalaen samtidigt er af en type, der er ækvidistant (modsat FP) kan man beregne progression som simpel difference mellem to scorer.
5. DNT giver (modsat FP og ILSA) altså mulighed for at se på elevens progression i perioden, hvilket er et langt bedre mål for skoleeffekt end blot det faglige niveau (et øjebliksbillede, ligesom det, der også måles med FP og ILSA). Både på individuelt (elev-niveau) og aggregeret (fx klasse, skole, kommune og national opgørelse) niveau.

Resultater fra ILSA, FP, og DNT

ILSA er komparative stikprøvebaserede undersøgelser, der skal kunne bruges til at sammenligne landes uddannelsessystemer og deres outcome til brug for forskning og policy-formulering. Der bliver derfor også kort tid efter en undersøgelse gjort et omfattende statistisk materiale offentligt tilgængeligt: Men de kan kun opgøres på nationalt niveau, da det er en stikprøve på landsplan kan fx ingen skole- eller kommuneresultater opgøres på forsvarlig måde.

FP udgør en serie af afsluttende eksaminationer af eleverne i folkeskolen. FP skal afprøve alle dele af faget hvorfor der ikke er noget problem i at undervisningen har fokus mod at eleverne kan klare sig bedst muligt i prøven (wash-back-effekter/Teaching-to-the-test er altovervejende positive – FP bruges som en

¹ ILSA er kort for "International Large Scale Assessment" hvilket dækker over undersøgelser som PISA, PIRLS, TIMMS, PIAAC og IEA.

pædagogisk løftestang af skolerne). FP har indtil for nylig været mest af ceremoniel betydning men har i den senere tid i stigende omfang fået tillagt en praktisk betydning, eftersom FP bruges som ved afgørelse af adgang til ungdomsuddannelse. Resultaterne rapporteres til elever (FP-resultat) og diverse registre, og resultaterne aggregeret til skoleniveau og landsniveau offentliggøres af Undervisningsministeriet.

DNT gennemføres obligatorisk 10 gange i løbet af folkeforløbet mellem 2. og 8. kl., heraf 4 og 2 gange for hhv. læsning og matematik. Der testes smalle og Uni-dimensionelle konstrukter (3 profilområder pr. test), og formålet er formativt, dvs. at lærerne skal blive klogere på hvad eleverne kan, med henblik på feedback og tilrettelæggelse af undervisningen. Hvis undervisningen blev orienteret mod testene ville det skævvride fokus for undervisningen (negativ washback). For at undgå fokus på summativ anvendelse/konkurrence mellem skoler, blev det ved lov forbudt at offentliggøre resultater, der kunne danne afsæt for ranking af resultater på noget niveau (fx elever, klasser, skoler, kommuner og regioner/landsdele).

Ved testenes lancering i 2010 blev alle skalaer normeret på percentil-niveau, og percentilresultater refererer derfor stadig til resultatfordeling i 2010. Det er en klar fordel at alle resultater er sammenlignelige over tid: Et resultat på "43" fx i tekstforståelse repræsenterer den samme dygtighed i 2010 som 2017. Imidlertid udgør det et informationsmæssigt problem for de skoler, der arbejder med progression (fx gennem Beregneren) at udviklingen i fordelingen ikke bliver offentliggjort i raschscorer, men kun i percentiler (jf [de nationale præstationsprofiler](#)) især da det fremgik, at resultaterne steg markant de først par år, sandsynligvis fordi skolerne blev bedre til at gennemføre testene så elevresultaterne blev mere retvisende.

Metode og datagrundlag bag ny reference for DNT

Der mangler altså en reference hvad angår progression på landsplan. Datamaterialet til en sådan er imidlertid blevet tilvejebragt i forbindelse med et projekt gennemført af EVA, hvor det faglige niveau i DNT og progressionen heri skulle anvendes som forklarende variabel til en analyse effekten af uddannelsen til diplom i ledelse (DIL). Denne nationale reference er blevet til i samarbejde mellem Danmarks Evalueringsinstitut (EVA) og NordicMetrics (NM) som i samarbejde har fremstillet et overblik over elevernes læring i matematik og dansk, læsning, målt ved de nationale test 2009/10 til 2015/16. Resultaterne af dette arbejde er blevet anvendt som forklarende variable i EVA's projekt "Effekt af lederuddannelse for skoleledere" (EVA 2017)².

Om det fremstillede datamateriale er det i rapporten anført:

- "Effekten på elevernes færdigheder i læsning og matematik undersøges ved hjælp af data fra de nationale test. Der kigges på udviklingen mellem læsetestene i 2., 4., 6. og 8. klasse og udviklingen mellem matematiktestene i 3. klasse og 6. klasse for de enkelte elever. De anvendte skalaer angiver elevernes læseniveau og matematikniveau baseret på deres Raschscorer i De Nationale Test.

² EVA (2017), Effekten af lederuddannelse af skoleledere, Dansk evalueringsinstitut, ISBN: (www) 978-87-7182-011-9 (Hentet 4. juni 2017 fra: <https://www.eva.dk/projekter/2015/effekt-af-lederuddannelse-for-skoleledere/hent-udgivelse/effekten-af-lederuddannelse-af-skoleledere/download>)

Ækvivaleringen af skalaerne er foretaget af NordicMetrics på grundlag af i alt ca. 1.200 elevers systematiske besvarelser af forskellige nationale test. Ækvivaleringen blev foretaget med henblik på at kunne sammenligne resultater fra test af den samme elev i læsning og matematik på forskellige tidspunkter med hinanden på en absolut skala." (EVA 2017, s. 16).

Videre oplyses at:

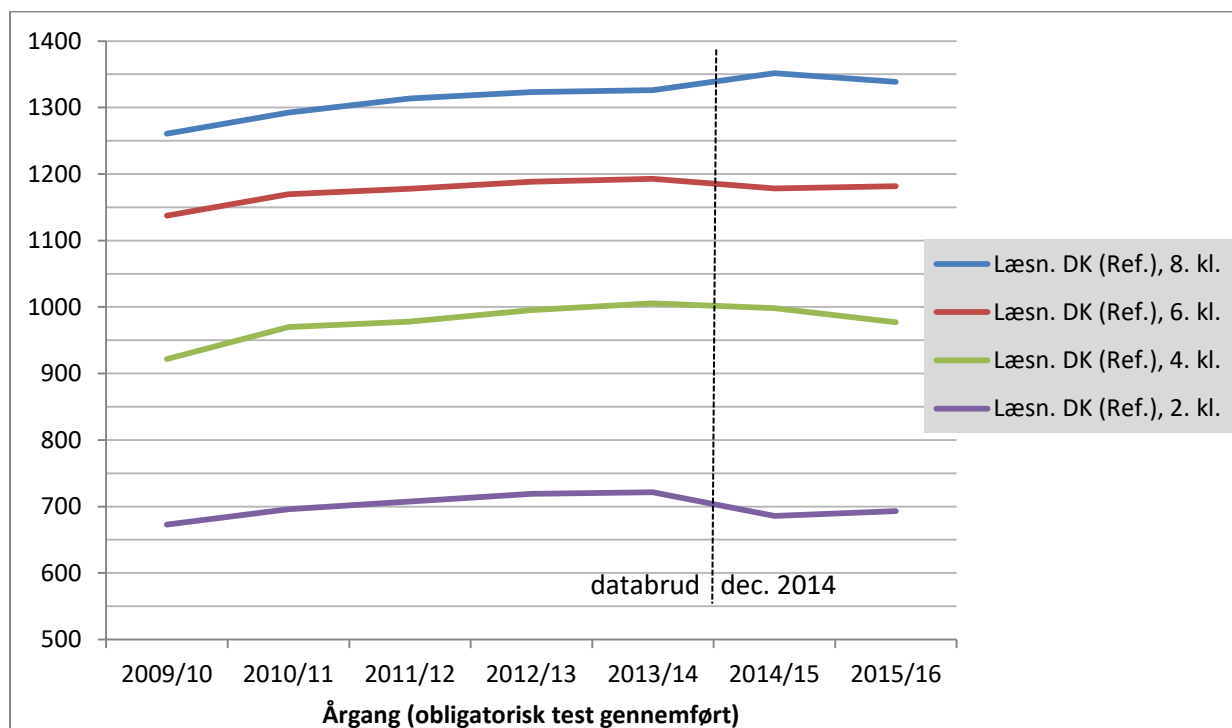
- "I undersøgelsen har vi afgrænset os til udelukkende at se på, hvad vi betegner som stabile institutioner. Ved dette forstår vi institutioner, som ikke har været flyttet eller gennemgået større sammenlægninger med mere i minimum fire år i tællingsperioden. En institution, som i 2009 flytter til en anden adresse, er således ikke medtaget i 2008 og 2009, men er med i 2010-14 (såfremt den selvfølgelig ikke flytter yderligere eller gennemgår større sammenlægninger). Vi danner her et datasæt med stabile skoler, som bruges sammen med de øvrige data. Dette datasæt er, hvad vi i det nedenstående betegner som det samlede institutionsregister." (EVA 2017, s. 36)

Af hensyn til stabiliteten i materialet blev altså valgt at fokusere på skoler, der ikke var under omstrukturering el.lign. i undersøgelsesperioden. Progression, der kan måles på disse skoler (med mere end halvdelen af landets elever - mellem godt 35 og 43 tusind elever pr årgang i perioden), vil på flere måder udgøre en solid reference for landets skoler/kommuner som helhed (det kan ellers være svært at vurdere hvor meget elevernes læring påvirkes af organisatoriske ændringer).

Fagligt niveau og udvikling fra 2009/10 til 2015/16

I figur 1 er vist den gennemsnitlige Rasch-score (transformeret som i beregneren) for alle årgange, der har deltaget i DNT, Dansk, læsning.

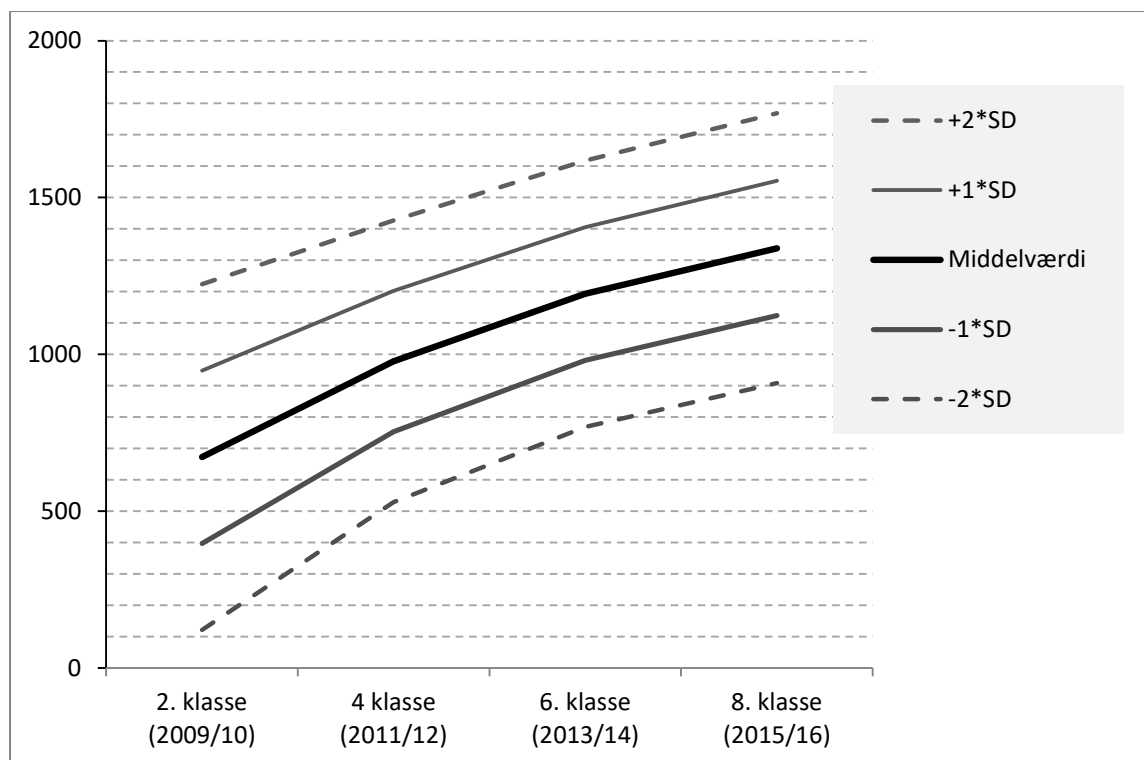
Figur 1 Fagligt niveau (tekstforståelse) for alle årgange, der har deltaget i DNT, Læsning.



Det fremgår at der det første par år var et svagt stigende niveau for alle test i direkte forlængelse af implementeringen, hvorefter udviklingen fladede ud. Dette passer med internationale erfaringer om at der i starten kan forekomme en implementeringsgevinst. Dette kan altså tilskrives skolernes bedre test anvendelse og instruktion af eleverne som alternativ forklaring til at eleverne var blevet dygtigere. I dec. 2014 blev alle items genafprøvet og ny itemparametre (dvs opgavesværhed) blev implementeret. Ministeriet forsøgte at minimere effekten af denne risiko for databrud (tidspunktet vist i figur 2. ved den lodrette stiplede linje) gennem tilpasning af percentilscore, men meddelte for en sikkerheds skyld, at det ikke var sikkert at testscorer før og efter december 2014 var helt sammenlignelige. Det fremgår at der kan være tale om et mindre skred (knæk på kurverne, størst for 2 og 8. klasse, men anslået på under 50 point), men at det altså er lykkedes at imødegå risikoen for markante databrud.

Perioden omfatter bl.a. den første årgang elever der har taget alle test (dem, der gik i 2. kl. i 2009/10 og som gik i 8. kl. 2015/16). I figur 1 er vist den gennemsnitlige progression ("Middelværdi") for denne årgang.

Figur 2 Progression (tekstforståelse) for den første årgang, der har taget alle 4 læsetest.



Endvidere er vist spredningen på elevernes dygtighed (over og under middelværdien er vist 1 og 2 gang standard-deviationen – det ville svare til hhv. 98-, 84-, 16- og 2-percentilen hvis testscorerne var normalfordelte, hvad de dog ikke helt er).

Ny reference for DNT

Vil man vurdere udvikling på aggregeret niveau (fx skole eller kommuneniveau) er der altså flere forhold der bør tages i betragtning (implementeringsgevinst, databrud). Til det formål kan anvendes de (tilnærmede)

landsresultater fra EVA's undersøgelse (EVA 2017), hvor der for både læsning og matematik blev beregnet middelværdi (Gennemsnit/Mean) og spredningen i data (standardafvigelsen/SD) for alle test i læsning og matematik, for alle år og for både fagligt niveau og progression.

Oplysninger om fagligt niveau og progression³ i læsning og matematik⁴, 2010-2016

Læsning, Absolut niveau	Fagligt niveau, 2. kl.			Fagligt niveau, 4. kl.			Fagligt niveau, 6. kl.			Fagligt niveau, 8. kl.		
	N Obs	Mean	SD	N Obs	Mean	SD	N Obs	Mean	SD	N Obs	Mean	SD
2009/10	35686	673	276	36254	922	230	37361	1137	193	35441	1261	215
2010/11	40169	696	265	41520	970	222	40779	1170	203	39202	1293	218
2011/12	42597	707	260	41943	978	224	41849	1178	208	40517	1314	217
2012/13	43262	719	253	41378	995	223	42709	1189	208	39442	1323	216
2013/14	42588	721	256	42884	1006	225	41896	1193	212	39884	1326	220
2014/15	43354	686	248	43034	998	237	40726	1179	197	39785	1351	216
2015/16	42003	693	256	41875	977	230	41591	1182	193	38477	1339	215

Læsning, Progression	Faglig progression læsning, fra 2. til 4. kl.			Faglig progression læsning, fra 4. til 6. kl.			Faglig progression læsning, fra 6. til 8. kl.		
	N Obs	Mean	SD	N Obs	Mean	SD	N Obs	Mean	SD
2009/10	-	-	-	-	-	-	-	-	-
2010/11	-	-	-	-	-	-	-	-	-
2011/12	33371	154,9	115,5	33610	131,6	90,5	31754	90,5	83,5
2012/13	37501	144,2	102,9	38436	105,5	83,1	34262	75,0	80,5
2013/14	39956	148,9	105,7	38814	108,1	87,0	35603	74,1	83,9
2014/15	40284	146,5	110,9	37980	96,5	89,7	35957	87,1	89,1
2015/16	39138	129,0	109,6	38918	89,4	88,6	34729	74,9	86,7

Matematik, Absolut niveau	Fagligt niveau 3. kl., matematik samlet			Fagligt niveau 6. kl., matematik samlet		
	N Obs	Mean	SD	N Obs	Mean	SD
2009/10	37544	732	163	37458	977	136
2010/11	40331	735	166	40599	1003	147
2011/12	41311	733	167	41829	1005	153
2012/13	42589	740	169	42503	1012	160
2013/14	43455	743	171	41793	1017	167
2014/15	42543	785	169	40767	1026	147
2015/16	43035	766	174	41545	1022	145

Matematik, Progression	Faglig progression i matematik, fra 3. til 6. kl.		
	N Obs	Mean	SD
2009/10	-	-	-
2010/11	-	-	-
2011/12	-	-	-
2012/13	33842	92,0	46,1
2013/14	36418	94,3	47,2
2014/15	37008	97,8	44,2
2015/16	37825	97,0	45,5

³ **Note 1:** "Læsning" betegner her scoren i profilområdet tekstforståelse, og "Matematik" betegner her beregnet samlet score, sammenvejet af de tre profilområders rasch-scorer (matematikscoren er altså ikke en ægte rasch-score).

⁴ **Note 2:** Progression er beregnet som standardiserede årsprogression (kun personer med testresultater "i begge ender" er medregnet, og for at gøre progressionerne sammenlignelige det på grundlag af antallet af dage mellem testene beregnet hvor meget eleven ville have gået frem på 365 dage)