



**MINISTERIET FOR
BØRN, UNDERVISNING
OG LIGESTILLING**
STYRELSEN
FOR IT OG LÆRING

Olof Palmes Allé 38
8200 Aarhus N
Tlf.nr.: 35 87 88 89
E-mail: stil@stil.dk
www.stil.dk
CVR-nr.: 13223459

Undersøgelse af de nationale tests reliabilitet

26.02.2016

Sammenfatning

I efteråret 2014 blev der i alt gennemført ca. 485.000 frivillige nationale tests. 296.000 elever deltog i de frivillige test, heraf deltog 32.600 elever i den samme test to gange.

I alle profilområder er der en statistisk signifikant positiv sammenhæng mellem elevdygtigheden bestemt ved første og ved andet forsøg i de frivillige test. Specielt i afkodning og tekstforståelse i dansk læsning samt i engelsk er der en høj korrelation mellem to gentagne test.

Generelt er korrelationen mellem elevens samlede vurdering i forsøg 1 og forsøg 2 på 0,79.

Eleverne opnår et statistisk signifikant bedre testresultat i deres andet forsøg i dansk læsning, matematik og engelsk.

I gennemsnit opnår eleverne 3-6 point mere i deres andet forsøg i dansk, læsning og matematik og 2 point mere i engelsk, fysik/kemi, biologi og geografi.

Forskellen i elevdygtigheden mellem første og andet forsøg bliver statistisk signifikant større des længere tid, der går mellem de to forsøg i de fleste profilområder i dansk læsning i 2., 4. og 6. klasse samt i matematik i 3. og i 6. klasse.

Ændringen i elevdygtigheden mellem første og andet forsøg bliver statistisk signifikant større des større usikkerheden er på den estimerede elevdygtighed.

Der er en stærk statistisk signifikant positiv sammenhæng mellem testresultaterne i de obligatoriske nationale test og elevernes karakter i folke-

skolens prøve i 9. klasse. Specielt tekstforståelse i 8. klasse korrelerer højt med folkeskolens prøve i dansk læsning.

80 procent af de elever, der samlet opnår vurderingen 'mangelfuld' i dansk læsning 8. klasse, får højst karakteren 4 i dansk læsning i folkeskolens prøve i 9. klasse. 80 procent af de elever, der samlet opnår vurderingen 'rigtig god' i dansk læsning 8. klasse, får mindst 7 i dansk læsning i folkeskolens prøve i 9. klasse. Næsten 60 procent af de elever, der samlet opnår vurderingen 'fremragende' i dansk læsning 8. klasse, får mindst 10 i dansk læsning i folkeskolens prøve i 9. klasse.

Tilsvarende er der en statistisk god sammenhæng mellem de kriteriebase-rede tilbagemeldinger i matematik i 6. klasse og folkeskolens prøve i fær-dighedsregning i 9. klasse. Fx opnår 75 procent af de elever, der samlet opnår vurderingen 'fremragende' i matematik 6. klasse, mindst 10 i fær-dighedsregning i folkeskolens prøve i 9. klasse.

Ved afholdelse af de frivillige test i efteråret 2014 fik de elever, der deltog to gange, i gennemsnit 25 procent af de samme opgaver anden gang som de fik første gang.

Des større forskel i den beregnede elevdygtighed mellem de to frivillige forsøg, des mindre er sandsynligheden for at eleven får de samme opgaver begge gange.

Når der er taget højde for elevens dygtighed bestemt ved første forsøg, er der generelt ingen statistisk signifikant sammenhæng mellem elevens dygtighed i andet forsøg og graden af gentestning. Det kan således ikke påvises, at genanvendelse af opgaver øger sandsynligheden for et bedre testresultat.

Baggrund

Skoler kan deltage i de frivillige nationale test hvert efterår. Et stigende antal skoler vælger denne mulighed. Folkeskoler har mulighed for at lade deres elever deltage i en given test op til to gange ud over den obligatoriske nationale test. Skoler, der ikke er omfattet af lov om deltagelse i de obligatoriske test, kan lade deres elever deltage i de frivillige test op til tre gange.

Nogle skoler lader eleverne deltage i den samme frivillige test med få dages eller få ugers mellemrum. Dette giver mulighed for at undersøge, i hvilken udstrækning eleverne opnår samme testresultat i to på hinanden følgende test.

Dette notat belyser testenes reliabilitet ved at se på testresultaterne for de elever, der deltog i samme test to gange i efteråret 2014.

Endvidere undersøges, i hvilken udstrækning testresultaterne stemmer overens med de prøveresultater, eleverne opnår i 9. klasse.

Endelig beskriver notatet omfanget af gentestning af enkelte opgaver, dvs. hvor mange opgaver får den enkelte elev i test nummer to, som eleven også fik i test nummer et.

Deltagelse i de frivillige nationale test efteråret 2014

Tabel 1. Antallet af afholdte frivillige nationale tests efteråret 2014

Test	Antal test	Antal elever	Heraf antal elever med to test
Dansk/læsning 2. klasse	62.765	56.708	6.057
Dansk/læsning 4. klasse	64.327	59.891	4.421
Dansk/læsning 6. klasse	61.126	55.964	5.134
Dansk/læsning 8. klasse	42.220	39.610	2.558
Matematik 3. klasse	61.027	53.427	7.590
Matematik 6. klasse	58.642	53.179	5.440
Engelsk 7. klasse	41.564	39.103	2.424
Fysik/kemi 8. klasse	29.264	28.056	1.169
Biologi 8. klasse	29.304	28.522	761
Geografi 8. klasse	31.032	29.964	1.028
Dansk som andetsprog 5.	2.415	2.366	49
Dansk som andetsprog 7.	1.839	1.815	24
I alt	485.525	295.744	32.566

I efteråret 2014 blev der gennemført godt 485.000 frivillige nationale tests. I alt deltog næsten 296.000 elever i de frivillige test, heraf deltog 32.566 (11 procent) elever i den samme test to gange. 1.514 institutioner deltog i de frivillige test, hvoraf 723 institutioner anvendte gentestning i mindst ét fag.

I gennemsnit gik der 22 dage (range: 0-53 dage) mellem første og anden test.

Der er relativt få elever, der tager testen i dansk som andetsprog mere end én gang, hvorfor denne test ikke behandles yderligere i dette notat.

Forskellen i elevdygtigheden mellem to gentagne frivillige test

Eleverne opnår et statistisk signifikant bedre testresultat i deres andet forsøg i dansk læsning, matematik og engelsk.

Tabel 2. Forskel mellem forsøg 1 og forsøg 2 i elevdygtigheden

Test	Profil-område	Antal elever	Gennemsnit (logit skala)	SD (logit skala)	Gennemsnit (percentil skala)
Dansk/læsning 2. klasse	1	6.057	0,17*	0,99	5
	2	6.057	0,30*	1,12	5
	3	6.057	0,27*	1,16	4
Dansk/læsning 4. klasse	1	4.421	0,15*	0,92	4
	2	4.421	0,22*	0,87	5
	3	4.421	0,18*	0,85	4
Dansk/læsning 6. klasse	1	5.134	0,18*	0,93	5
	2	5.134	0,30*	0,83	6
	3	5.134	0,12*	0,81	3
Dansk/læsning 8. klasse	1	2.558	0,23*	0,98	5
	2	2.558	0,27*	1,07	6
	3	2.558	0,07*	0,75	1
Matematik 3. klasse	1	7.590	0,16*	0,88	4
	2	7.590	0,18*	0,89	6
	3	7.590	0,21*	1,02	5
Matematik 6. klasse	1	5.440	0,16*	0,89	5
	2	5.440	0,13*	0,80	5
	3	5.440	0,16*	0,90	5
Engelsk 7. klasse	1	2.424	0,06*	0,64	1
	2	2.424	0,14*	0,86	3
	3	2.424	0,14*	0,72	3
Fysik/kemi 8. klasse	1	1.169	0,04	0,71	1
	2	1.169	0,04	0,69	2
	3	1.169	0,04	0,71	1
Biologi 8. klasse	1	761	0,08*	0,72	3
	2	761	0,01	0,65	1
	3	761	0,03	0,67	2
Geografi 8. klasse	1	1.028	0,08*	0,65	3
	2	1.028	0,04	0,68	2
	3	1.028	0,03	0,70	1

* Statistisk signifikant forskellig fra 0

I gennemsnit opnår eleverne 3-6 point mere i deres andet forsøg i dansk, læsning og matematik og 2 point mere i engelsk, fysik/kemi, biologi og geografi (tabel 2).

Forskellen i elevdygtigheden mellem første og andet forsøg bliver statistisk signifikant større des længere tid, der går mellem de to forsøg, for dansk læsning i 2. klasse i alle tre profilområder, for dansk læsning i 4. klasse for to af profilområderne, for dansk læsning i 6. klasse i ét profilområde og for matematik i 3. og i 6. klasse i to af profilområderne (tabel 3).

Tabel 3. Forskel mellem forsøg 1 og forsøg 2 i elevdygtigheden målt på percentil skalaen i forhold til tiden mellem de to forsøg

Test	Profil- område	Højest 1 uge		1 til 2 uger		2-3 uger		3+ uger	
		Antal elever	Gns.	Antal elever	Gns.	Antal elever	Gns.	Antal elever	Gns.
Dansk/læsning 2. klasse	1*	1.345	2	1.275	4	1.086	6	2.351	6
	2*	1.345	2	1.275	4	1.086	5	2.351	6
	3*	1.345	2	1.275	3	1.086	5	2.351	6
Dansk/læsning 4. klasse	1	637	1	840	3	852	5	2.092	4
	2*	637	3	840	5	852	4	2.092	5
	3*	637	2	840	3	852	5	2.092	5
Dansk/læsning 6. klasse	1	1.132	3	730	5	801	5	2.471	5
	2*	1.132	5	730	6	801	7	2.471	7
	3	1.132	2	730	3	801	4	2.471	3
Dansk/læsning 8. klasse	1	433	6	571	6	272	4	1.282	5
	2	433	6	571	7	272	8	1.282	6
	3	433	-2	571	2	272	6	1.282	1
Matematik 3. klasse	1*	1.654	3	1.313	5	1.277	3	3.346	6
	2	1.654	6	1.313	6	1.277	5	3.346	6
	3*	1.654	4	1.313	5	1.277	3	3.346	6
Matematik 6. klasse	1*	876	1	832	5	1.058	6	2.674	6
	2	876	2	832	5	1.058	6	2.674	5
	3*	876	0	832	4	1.058	6	2.674	6
Engelsk 7. klasse	1	400	-1	346	2	363	2	1.315	2
	2	400	2	346	3	363	2	1.315	3
	3	400	2	346	3	363	3	1.315	3
Fysik/kemi 8. klasse	1	344	-0	362	2	119	1	344	2
	2	344	-1	362	4	119	8	344	2
	3	344	-2	362	4	119	4	344	2
Biologi 8. klasse	1	95	3	200	2	52	14	414	3
	2	95	3	200	0	52	-1	414	1
	3	95	3	200	-1	52	3	414	2
Geografi 8. klasse	1	138	3	145	1	293	4	452	4
	2	138	-1	145	1	293	5	452	1
	3	138	-0	145	-2	293	4	452	1

* Statistisk signifikant udvikling over tid. Den statistiske analyse er foretaget på logit skalaen

Forskellen i elevdygtigheden mellem første og andet forsøg bliver statistisk signifikant større des større usikkerheden, SEM, er på den estimerede elevdygtighed. Specielt giver en stor usikkerhed på elevdygtigheden i forsøg 2 en større forskel i elevdygtighederne mellem de to forsøg.

Hvis usikkerheden kommer under 0,40 i begge forsøg, da vil 50 procent af eleverne højst opleve udsving på ± 10 point omkring den observerede gennemsnitlige forskel mellem forsøg 1 og 2 i dansk og matematik i samtlige profilområder.

Tilsvarende afhænger forskellen i elevdygtigheden mellem første og andet forsøg statistisk signifikant af det antal opgaver, der løses. Des flere opgaver eleven når at besvare, des mindre forskel i elevdygtigheden mellem de to forsøg.

Endelig afhænger forskellen i elevdygtigheden mellem første og andet forsøg statistisk signifikant af forskellen i det antal opgaver, der løses i de to forsøg. Des større forskel, der er i det antal opgaver eleven når at besvare i de to forsøg, des større forskel i elevdygtigheden mellem de to forsøg

Tabel 4. Forskel mellem første og andet forsøg i elevdygtigheden målt på fem-trins skalaen. Andel elever (%)

Test	Profil-område	Antal elever	Falder mere	Falder 1 trin	Ingen ændring	Stiger 1 trin	Stiger mere
Dansk/læsning 2. klasse	1	6.057	6	17	42	26	9
	2	6.057	2	13	56	26	4
	3	6.057	2	14	53	26	5
	Samlet	6.057	1	12	58	26	2
Dansk/læsning 4. klasse	1	4.421	5	18	43	27	7
	2	4.421	2	14	52	27	5
	3	4.421	3	16	50	26	5
	Samlet	4.421	1	12	61	25	2
Dansk/læsning 6. klasse	1	5.134	5	19	42	25	10
	2	5.134	1	12	53	30	4
	3	5.134	4	17	48	24	7
	Samlet	5.134	0	12	60	26	2
Dansk/læsning 8. klasse	1	2.558	6	16	45	23	11
	2	2.558	3	14	48	28	8
	3	2.558	4	18	48	25	4
	Samlet	2.558	0	13	59	26	2
Matematik 3. klasse	1	7.590	5	18	41	26	10
	2	7.590	7	17	39	27	12
	3	7.590	4	17	44	26	8
	Samlet	7.590	1	14	52	28	4
Matematik 6. klasse	1	5.440	6	19	39	26	11
	2	5.440	7	18	37	25	12
	3	5.440	6	19	39	26	10
	Samlet	5.440	1	15	52	29	3
Engelsk 7. klasse	1	2.424	2	21	49	24	4
	2	2.424	4	17	51	25	5
	3	2.424	1	15	56	26	2
	Samlet	2.424	0	12	68	20	0
Fysik/kemi 8. klasse	1	1.169	12	19	34	23	13
	2	1.169	11	20	33	22	14
	3	1.169	13	20	30	23	14
	Samlet	1.169	3	20	47	26	4
Biologi 8. klasse	1	761	9	21	35	20	16
	2	761	11	22	33	22	11
	3	761	9	20	37	23	11
	Samlet	761	2	22	50	22	4
Geografi 8. klasse	1	1.028	8	18	38	26	10
	2	1.028	9	20	36	25	10
	3	1.028	9	23	33	25	10
	Samlet	1.028	1	21	52	24	2

I dansk læsning er der højst 3 procent af testforløbene, hvor eleven oplever en ændring på mere end et trin på fem-trins skalaen i den samlede vurdering (tabel 4). I profilområde1, sprogforståelse, er der ca. 15 procent af testforløbene, hvor der sker en ændring på mere end et trin på fem-trins skalaen.

I matematik er der højst 5 procent af testforløbene, hvor eleven oplever en ændring på mere end et trin på fem-trins skalaen i den samlede vurdering. I profilområde2, geometri, er der ca. 19 procent af testforløbene, hvor der sker en ændring på mere end et trin på fem-trins skalaen.

For hver elev beregnes en standardiseret forskel, U , på den estimerede elevdygtighed

$$U = \frac{D_1 - D_2}{\sqrt{(SEM_1^2 + SEM_2^2)}}$$

Her er D_1 og D_2 lig den estimerede elevdygtighed til første og andet forsøg, mens SEM_1 og SEM_2 er de tilhørende estimerede usikkerheder. Justeres endvidere med den gennemsnitlige niveauforskel fra første til andet forsøg, da vil U følge en standard normalfordeling. I denne fordeling forventes 95 % af elevernes resultater at ligge mellem $-1,96$ og $+1,96$ mens 5 % af elevernes resultater forventes at ligge udenfor $\pm 1,96$.

Tabel 5 Andelen af elever, hvor forskellen i elevdygtigheden i første og andet forsøg ligger uden for 95 % sikkerhedsinterval

Test	Antal elever	Profilområde 1	Profilområde 2	Profilområde 3
Dansk/læsning 2. klasse	6.057	16 %	20 %	24 %
Dansk/læsning 4. klasse	4.421	13 %	11 %	12 %
Dansk/læsning 6. klasse	5.134	11 %	8 %	11 %
Dansk/læsning 8. klasse	2.558	14 %	14 %	8 %
Matematik 3. klasse	7.590	12 %	11 %	14 %
Matematik 6. klasse	5.440	13 %	9 %	12 %
Engelsk 7. klasse	2.424	5 %	12 %	9 %
Fysik/kemi 8. klasse	1.169	8 %	6 %	9 %
Biologi 8. klasse	761	8 %	5 %	6 %
Geografi 8. klasse	1.028	7 %	7 %	8 %

I alt ligger 12 procent af forskellene mellem elevdygtighederne i forsøg 1 og forsøg 2 udenfor det forventede, hvilket er lidt mere end de 5 procent, der forventes i en normalfordeling (tabel 5). Spredningen i elevernes resultater mellem første og andet forsøg er således lidt større end, der forventes, og større end usikkerheden (SEM) på elevdygtighederne kan forklare. Specielt i dansk læsning i 2. klasse er der en større variation

mellem elevernes resultater i første og andet forsøg end, der umiddelbart kan forklares ved usikkerheden i estimeringen af elevdygtighederne.

I engelsk og i de tre naturfag er der generelt en tilfredsstillende overensstemmelse mellem den estimerede elevdygtighed i første og andet forsøg.

Sammenhæng mellem to gentagne test

Sammenhængen mellem elevernes to forsøg i de frivillige tests udtrykt ved korrelationskoefficienten ses af tabel 6. Denne korrelation kaldes også i nogle sammenhænge for 'test-retest-reliability coefficient' eller 'coefficient of stability'. Korrelationskoefficienten udtrykker graden af sammenhæng mellem to målinger og ligger mellem ± 1 og $+1$.

Tabel 6 Sammenhængen mellem forsøg 1 og forsøg 2 i elevdygtigheden målt på logit skalaen. Pearson korrelationen

Test	Antal elever	Profil-område 1	Profil-område 2	Profil-område 3
Dansk/læsning 2. klasse	6.057	0,57*	0,80*	0,75*
Dansk/læsning 4. klasse	4.421	0,63*	0,78*	0,75*
Dansk/læsning 6. klasse	5.134	0,56*	0,81*	0,73*
Dansk/læsning 8. klasse	2.558	0,66*	0,74*	0,74*
Matematik 3. klasse	7.590	0,60*	0,56*	0,66*
Matematik 6. klasse	5.440	0,61*	0,56*	0,60*
Engelsk 7. klasse	2.424	0,76*	0,77*	0,80*
Fysik/kemi 8. klasse	1.169	0,45*	0,41*	0,39*
Biologi 8. klasse	761	0,41*	0,49*	0,50*
Geografi 8. klasse	1.028	0,47*	0,47*	0,45*

* Statistisk signifikant forskellig fra 0

I alle profilområder er der en statistisk signifikant positiv sammenhæng mellem elevdygtigheden bestemt ved første og ved andet frivillig test forsøg.

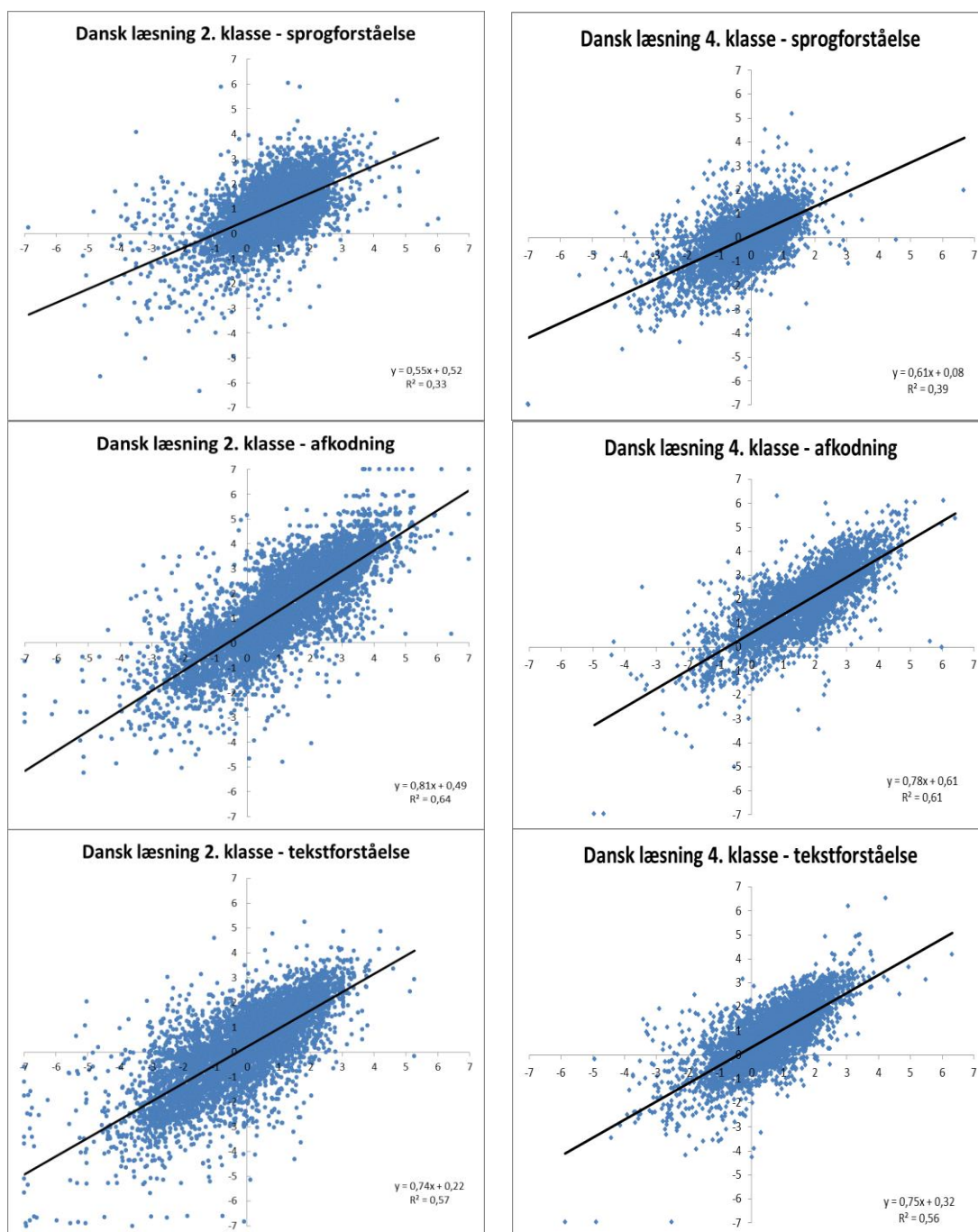
En korrelationskoefficient på 0,70 eller derover anses ofte som tilfredsstillende i forbindelse med måling af reliabiliteten¹. Disse er markeret med **fed** i tabel 6. Det drejer sig om afkodning (profilområde 2) og tekstforståelse (profilområde 3) i dansk læsning samt om alle tre profilområder i engelsk. En korrelationskoefficient under 0,60 anses som regel som utilfredsstillende. Hertil skal det dog bemærkes, at en korrelationskoefficient også afhænger af variationen i resultaterne mellem eleverne. Hvis variationen mellem eleverne er lille, vil korrelationen være svag uanset at elevernes testresultater i de to forsøg afspejler samme dygtighed. Variationen mellem elevernes resultater er relativ lille i geometri (profilområde 2 i matematik) samt i de tre naturfag i 8. klasse.

¹ Generel guideline til vurdering af reliabilitet: '0,0-0,5' = uacceptabel; '0,5-0,6' = dårlig; '0,6-0,7' = tvivlsom; '0,7-0,8' = acceptabel; '0,8-0,9' = god; '0,9-1,0' = fremragende

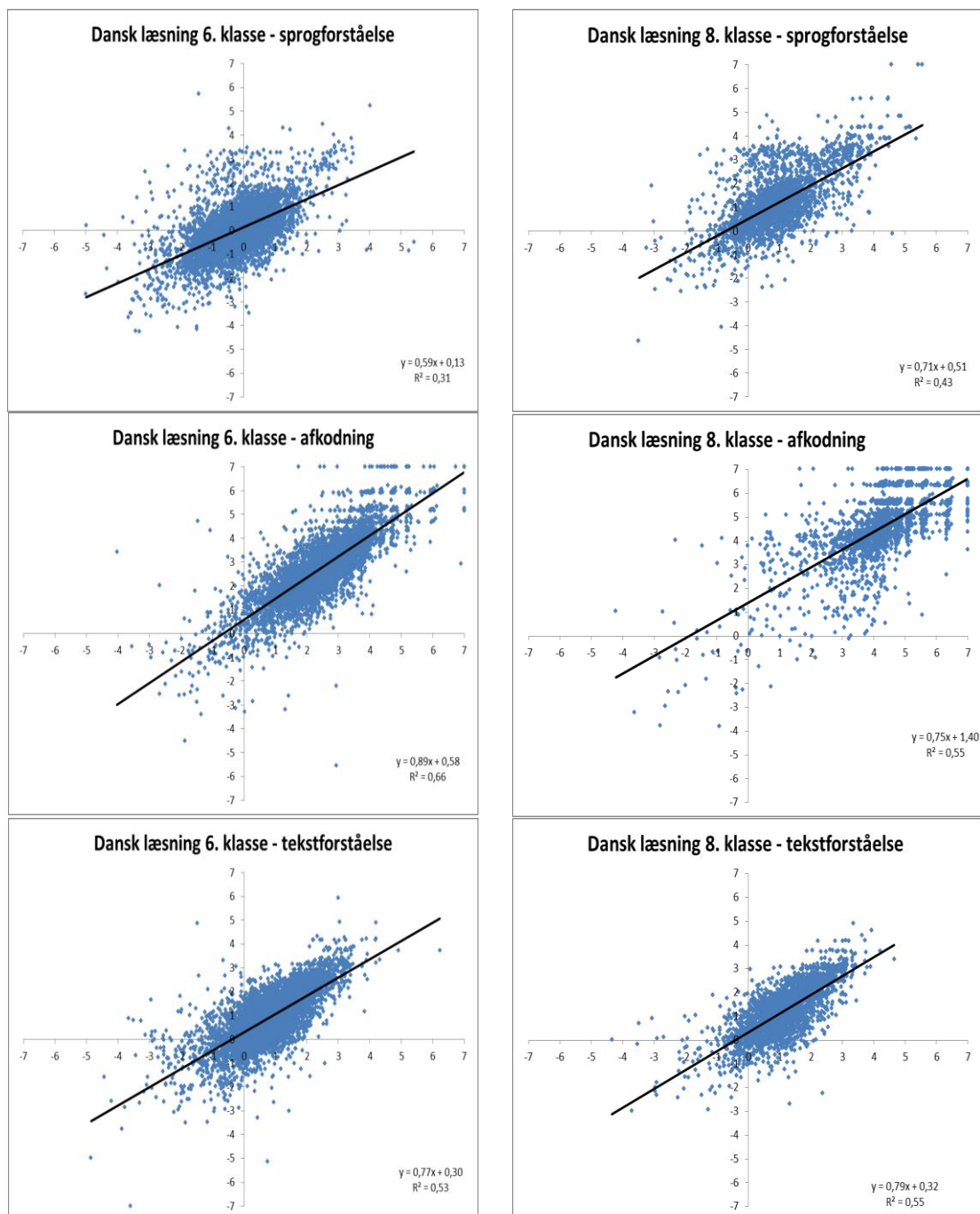
Korrelationen mellem elevens samlede vurdering i forsøg 1 og forsøg 2 er totalt set på 0,79.

I figur 1 – figur 5 ses sammenhængen i elevdygtigheden udtrykt på logit skalaen mellem elevernes to forsøg i de frivillige test.

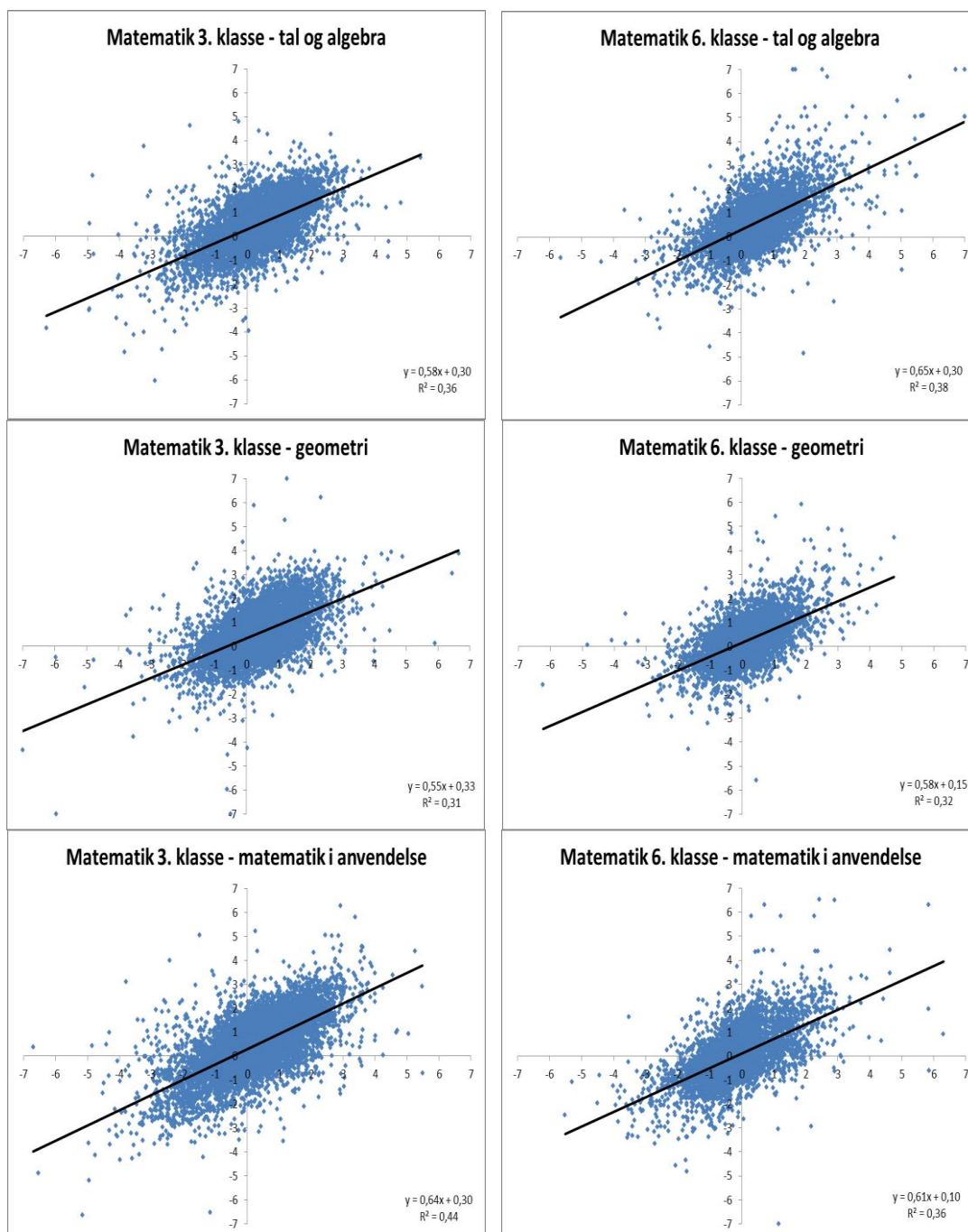
Figur 1. Sammenhængen mellem elevdygtigheden beregnet i første test (x-aksen) og anden test (y-aksen). Dansk læsning 2. og 4. klasse



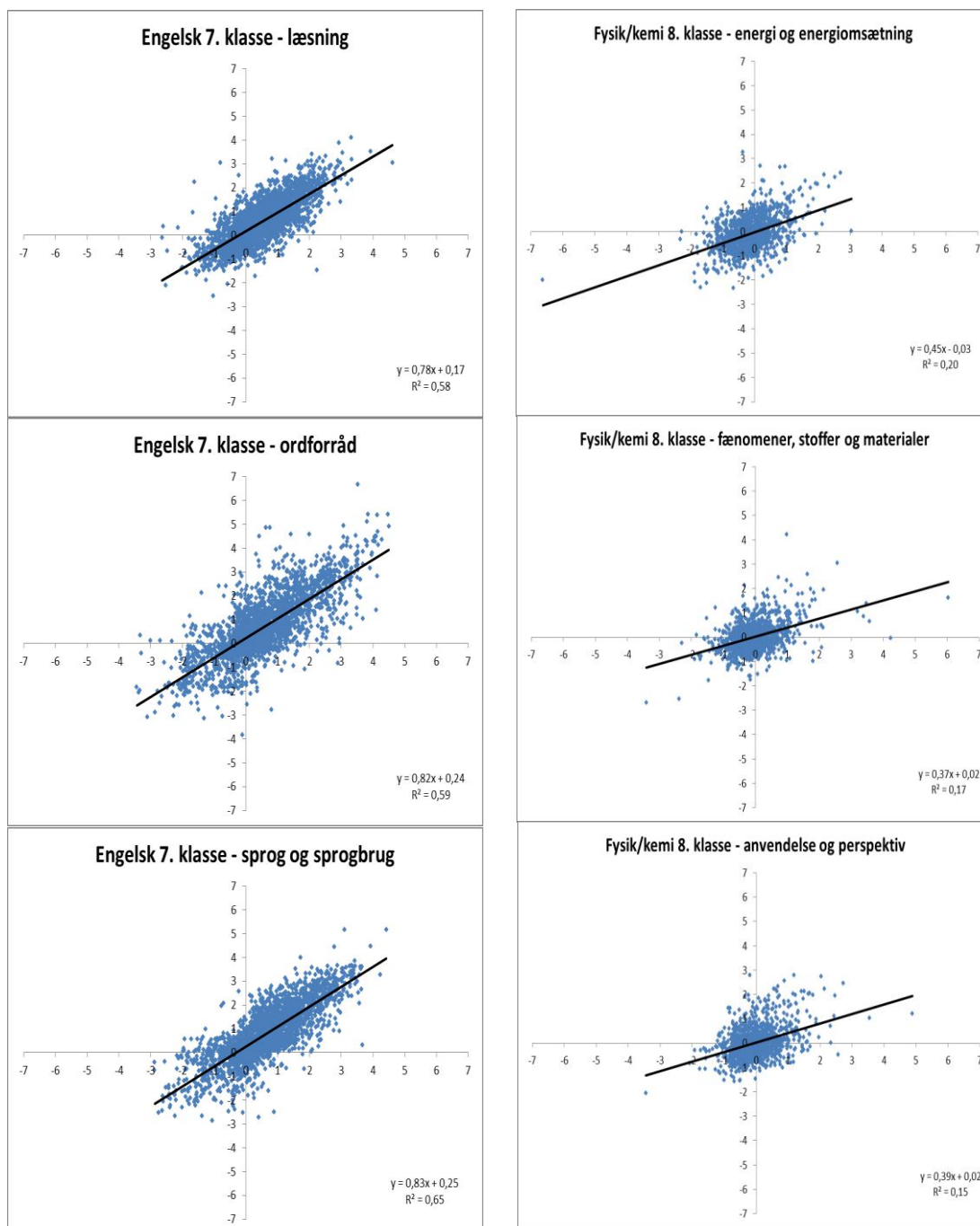
Figur 2. Sammenhængen mellem elevdygtigheden beregnet i første test (x-aksen) og anden test (y-aksen). Dansk læsning 6. og 8. klasse



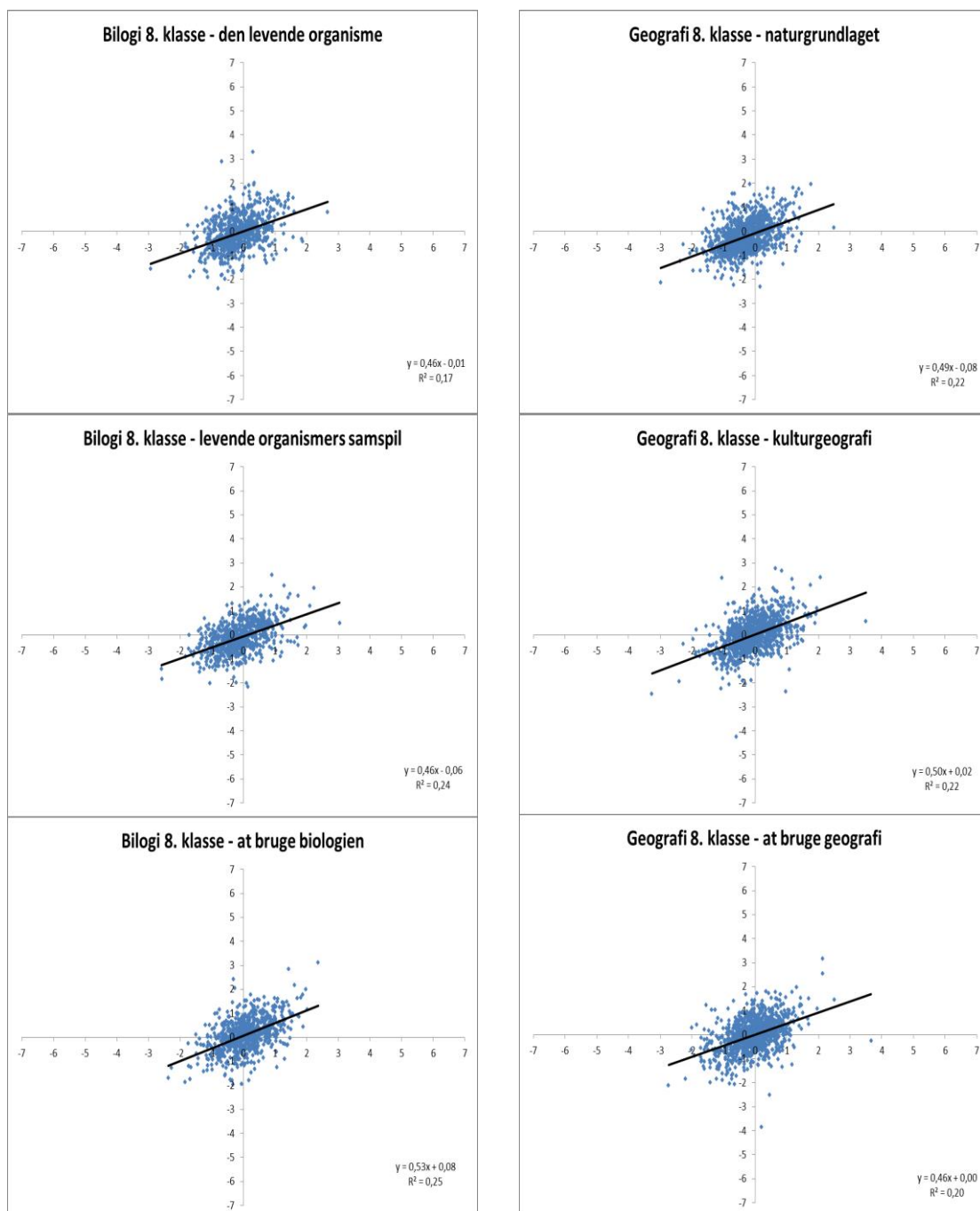
Figur 3. Sammenhængen mellem elevdygtigheden beregnet i første test (x-aksen) og anden test (y-aksen). Matematik 3. og 6. klasse



Figur 4. Sammenhængen mellem elevdygtigheden beregnet i første test (x-aksen) og anden test (y-aksen). Engelsk og fysik/kemi



Figur 5. Sammenhængen mellem elevdygtigheden beregnet i første test (x-aksen) og anden test (y-aksen). Biologi og geografi



Sammenhæng mellem elevernes resultater i de obligatoriske nationale test og folkeskolens prøver

Sammenhængen mellem elevernes karakterer i folkeskolens 9.-klasseprøver fra sommeren 2015 sammenlignes med elevernes testresultater i de nationale test. Det er kun elever med både en prøvekarakter og et testresultat, der indgår i analysen.

Dansk læsning

Elevernes karakter i dansk læsning i folkeskolens prøve i 9. klasse fra sommeren 2015 sammenholdes med elevernes testresultater i de obligatoriske nationale test i dansk læsning i 8. klasse fra foråret 2014.

Tabel 7 Sammenhængen mellem folkeskolens prøve i dansk læsning og testresultatet i dansk læsning i de obligatoriske nationale test på logit skalaen. Spearman korrelationen

Test	Antal elever	Sprogforståelse	Afkodning	Tekstforståelse
Dansk/læsning 8. klasse	48.382	0,44*	0,53*	0,62*

* Statistisk signifikant forskellig fra 0

Der er en statistisk signifikant positiv sammenhæng mellem testresultaterne i de obligatoriske nationale test i 8. klasse dansk læsning og elevernes karakter i folkeskolens prøve i dansk læsning året efter i 9. klasse. Specielt tekstforståelse korrelerer højt med folkeskolens prøve.

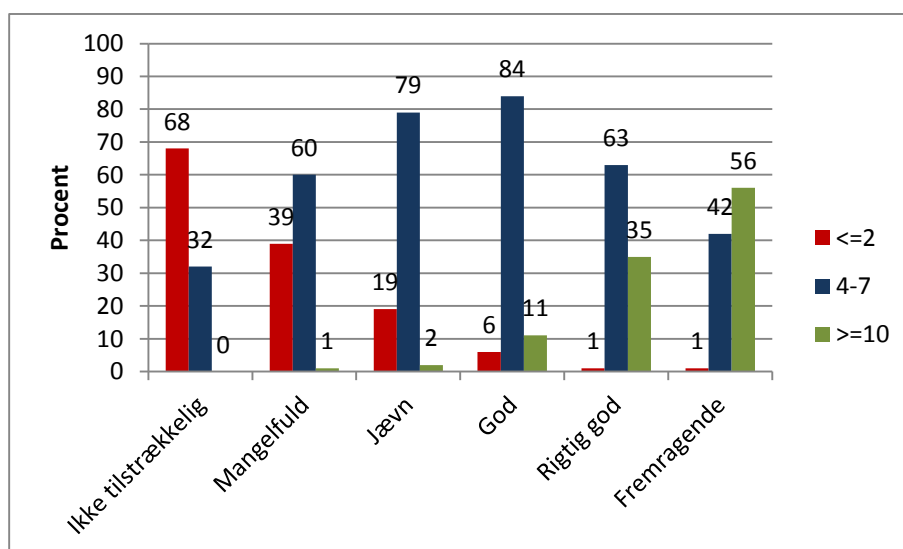
Af tabel 8 ses, at 86 procent af de elever, der ligger på 'Mangelfuld' i tekstforståelse, har fået karakteren 4 eller derunder i folkeskolens prøve i dansk læsning året efter, mens 81 procent af de elever, der ligger på 'Jævn' i tekstforståelse, har fået karakteren 4 eller 7 i folkeskolens prøve i dansk læsning året efter. 72 procent af de elever, der ligger på 'Fremragende' i tekstforståelse, har mindst fået karakteren 10 året efter.

Sammenhængen mellem den samlede vurdering i dansk læsning i 8. klasse i de obligatoriske nationale test og folkeskolens prøve i dansk læsning i 9. klasse ses i figur 6. 68 procent af de elever, der opnår en ikke tilstrækkelig præstation i de nationale test, får året efter højst karakteren 2 i folkeskolens prøve i dansk læsning. 84 procent af de elever, der opnår en god præstation i de nationale test, får året efter karakteren 4 eller 7 i folkeskolens prøve i dansk læsning. 56 procent af de elever, der opnår en fremragende præstation i de nationale test, får året efter mindst karakteren 10 i folkeskolens prøve i dansk læsning.

Tabel 8 Fordelingen på karakteren fra folkeskolens prøve i dansk læsning efter elevernes testresultat i de obligatoriske nationale test i dansk læsning 8. klasse udtrykt på den kriteriebaserede skala. Andel elever (%)

Profil-område	Testresultat	Karakter							I alt
		-3	0	2	4	7	10	12	
Sprog forståelse	Ikke tilstrækkelig	0	25	32	38	5	0	0	100
	Mangelfuld	0	7	18	56	18	1	0	100
	Jævn	0	3	9	46	36	6	1	100
	God	0	1	4	30	43	16	7	100
	Rigtig god	0	1	2	20	37	22	19	100
	Fremragende	0	1	3	27	40	16	13	100
Afkodning	Ikke tilstrækkelig	0	26	28	38	7	1	0	100
	Mangelfuld	0	11	22	50	15	2	0	100
	Jævn	0	4	15	53	24	3	1	100
	God	0	1	4	29	40	16	10	100
	Rigtig god								
	Fremragende								
Tekst forståelse	Ikke tilstrækkelig	0	25	29	38	6	0	0	100
	Mangelfuld	0	10	23	53	13	1	0	100
	Jævn	0	3	12	54	27	3	0	100
	God	0	1	5	38	44	10	3	100
	Rigtig god	0	0	1	16	44	25	14	100
	Fremragende	0	0	0	3	24	29	43	100
Samlet	Ikke tilstrækkelig	0	33	35	28	4	0	0	100
	Mangelfuld	0	13	26	51	9	1	0	100
	Jævn	0	4	15	56	23	2	0	100
	God	0	1	5	41	43	9	2	100
	Rigtig god	0	0	1	19	44	23	12	100
	Fremragende	0	0	1	10	32	25	31	100

Figur 6 De nationale test i dansk læsning 8. klasse og folkeskolens prøve i dansk læsning 9. klasse



Matematik

Elevernes karakter i matematiske færdigheder i folkeskolens prøve i 9. klasse fra sommeren 2015 sammenholdes med elevernes testresultater i de obligatoriske nationale test i matematik i 6. klasse fra foråret 2012.

Tabel 9 Sammenhængen mellem folkeskolens prøve i matematiske færdigheder og testresultatet i matematik i de obligatoriske nationale test på logit skalaen. Spearman korrelationen

Test	Antal elever	Tal og algebra	Geometri	Matematik i anvendelse
Matematik 6. klasse	52.345	0,41*	0,50*	0,51*

* Statistisk signifikant forskellig fra 0

Der er en statistisk signifikant positiv sammenhæng mellem testresultaterne i de obligatoriske nationale test i 6. klasse matematik og elevernes karakter i folkeskolens prøve i matematiske færdigheder tre år efter i 9. klasse. Korrelationen er ikke så stor som for dansk læsning, men i matematik er der også 3 år mellem de obligatoriske nationale test og folkeskolens prøve.

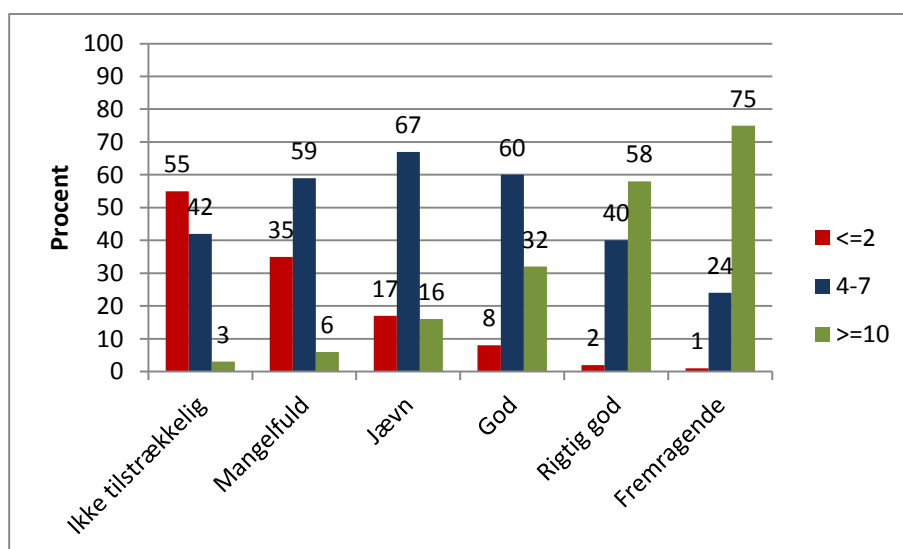
Af tabel 10 ses, at 63 procent af de elever, der ligger på 'Jævn' i tal og algebra, har fået karakteren 4 eller 7 i folkeskolens prøve i matematiske færdigheder tre år efter, mens 84 procent af de elever, der ligger på 'Fremragende' i tal og algebra, har fået karakteren 7 eller derover i folkeskolens prøve i matematik tre år efter.

Sammenhængen mellem den samlede vurdering i matematik i 6. klasse i de obligatoriske nationale test og folkeskolens prøve i færdighedsregning i 9. klasse ses i figur 7. 55 procent af de elever, der opnår en ikke tilstrækkelig præstation i de nationale test får tre år efter højst karakteren 2 i folkeskolens prøve i færdighedsregning. 58 procent af de elever, der opnår en rigtig god præstation og 75 procent af de elever, der opnår en fremragende præstation i de nationale test, får tre år efter mindst karakteren 10 i folkeskolens prøve i færdighedsregning.

Tabel 10 Fordelingen på karakteren fra folkeskolens prøve i matematiske færdigheder efter elevernes testresultat i de obligatoriske nationale test i matematik 6. klasse udtrykt på den kriteriebaserede skala. Andel elever (%)

Profil-område	Testresultat	Karakter							I alt
		-3	0	2	4	7	10	12	
Tal og algebra	Ikke tilstrækkelig	0	15	33	31	16	4	1	100
	Mangelfuld	0	5	23	35	26	9	2	100
	Jævn	0	2	12	29	34	17	6	100
	God	0	1	7	20	34	26	12	100
	Rigtig god	0	0	3	12	28	32	24	100
	Fremragende	0	0	4	12	27	28	29	100
Geometri	Ikke tilstrækkelig	0	13	36	31	15	4	1	100
	Mangelfuld	0	6	24	36	26	7	2	100
	Jævn	0	2	13	31	34	17	4	100
	God	0	0	4	15	32	30	19	100
	Rigtig god	0	1	.	13	34	30	21	100
	Fremragende	0	0	12	9	18	35	26	100
Matematik i anvendelse	Ikke tilstrækkelig	0	14	33	33	15	5	0	100
	Mangelfuld	0	5	24	37	26	7	1	100
	Jævn	0	2	15	31	33	15	3	100
	God	0	1	7	23	37	23	9	100
	Rigtig god	0	0	2	9	29	35	24	100
	Fremragende	0	0	0	3	13	35	49	100
Samlet	Ikke tilstrækkelig	0	18	37	30	12	3	0	100
	Mangelfuld	0	7	28	38	21	5	1	100
	Jævn	0	2	15	34	33	13	3	100
	God	0	1	7	23	37	24	8	100
	Rigtig god	0	0	2	10	30	35	23	100
	Fremragende	0	0	1	5	18	33	42	100

Figur 7 De nationale test i matematik 6. klasse og folkeskolens prøve i færdighedsregning i 9. klasse



Engelsk

Elevernes karakter i mundtlig engelsk i folkeskolens prøve i 9. klasse fra sommeren 2015 sammenholdes med elevernes testresultater i de obligatoriske nationale test i engelsk i 7. klasse fra foråret 2013.

Tabel 11 Sammenhængen mellem folkeskolens prøve i mundtlig engelsk og testresultatet i engelsk i de obligatoriske nationale test på logit skalaen. Spearman korrelationen

Test	Antal elever	Læsning	Ordfør-råd	Sprog og sprogbrug
Engelsk 7. klasse	47.553	0,60*	0,60*	0,63*

* Statistisk signifikant forskellig fra 0

Der er en statistisk signifikant positiv sammenhæng mellem testresultaterne i de obligatoriske nationale test i engelsk 7. klasse og elevernes karakter i folkeskolens prøve i mundtlig engelsk to år efter i 9. klasse.

Fysik/kemi, biologi og geografi

Elevernes karakter i fysik/kemi, biologi og geografi i folkeskolens prøve i 9. klasse fra sommeren 2015 sammenholdes med elevernes testresultater i de obligatoriske nationale test i 8. klasse fra foråret 2014.

Tabel 12 Sammenhængen mellem folkeskolens prøve i fysik/kemi, biologi og geografi og testresultatet i de obligatoriske nationale test på logit skalaen. Spearman korrelationen

Test	Antal elever	Profilområde 1	Profilområde 2	Profilområde 3
Fysik/kemi 8. klasse	46.522	0,40*	0,38*	0,38*
Biologi**	12.164	0,54*	0,57*	0,57*
Geografi**	11.991	0,51*	0,50*	0,55*

* Statistisk signifikant forskellig fra 0. ** Prøvefag til udtræk

Der er en statistisk signifikant positiv sammenhæng mellem testresultaterne i de obligatoriske nationale test i de tre naturfag og elevernes karakter i folkeskolens prøve året efter i 9. klasse.

Genanvendelse af testopgaver

Eleverne kan ikke få den samme opgave flere gange i ét og samme testforløb. Derimod kan eleverne godt blive præsenteret for nogle af de samme opgaver igen, hvis eleven tager den samme test flere gange. Hvis en elev tager den samme test som frivillig test to gange eller tager såvel den frivillige test som den tilsvarende obligatoriske test, da kan eleven anden gang godt blive præsenteret for nogle af de samme opgaver. Dette afsnit undersøger omfanget af denne form for gentestning.

Andelen af de opgaver eleven bliver præsenteret for anden gang, som eleven også har besvaret første gang i de frivillige test ses af tabel 13.

Ved afholdelse af de frivillige test i efteråret 2014 fik de elever, der deltog to gange, i gennemsnit 25 procent af de samme opgaver anden gang som de fik første gang.

Tabel 13 Andelen af de opgaver eleven får anden gang ved de frivillige test som eleven også fik første gang (%)

Test	Profil-område	Antal elever	25 % fraktil	Gennemsnit	75 % fraktil
Dansk/læsning 2. klasse	1	6.057	10	27	41
	2	6.057	8	30	47
	3	6.057	7	29	47
Dansk/læsning 4. klasse	1	4.421	6	18	25
	2	4.421	12	34	53
	3	4.421	7	29	47
Dansk/læsning 6. klasse	1	5.134	5	15	20
	2	5.134	30	53	76
	3	5.134	8	32	53
Dansk/læsning 8. klasse	1	2.558	10	28	40
	2	2.558	63	71	90
	3	2.558	13	42	71
Matematik 3. klasse	1	7.590	4	17	25
	2	7.590	6	17	25
	3	7.590	7	19	27
Matematik 6. klasse	1	5.440	0	12	18
	2	5.440	0	13	19
	3	5.440	4	13	20
Engelsk 7. klasse	1	2.424	7	26	41
	2	2.424	5	17	25
	3	2.424	14	31	41
Fysik/kemi 8. klasse	1	1.169	0	13	18
	2	1.169	0	12	18
	3	1.169	6	14	20
Biologi 8. klasse	1	761	5	14	20
	2	761	6	15	22
	3	761	6	16	23
Geografi 8. klasse	1	1.028	8	17	25
	2	1.028	4	13	19
	3	1.028	7	16	22
I alt	Samlet		11	25	35

Der er store forskelle på graden af gentestning mellem de enkelte profilområder. I dansk, læsning 8. klasse er gentestningsgraden på godt 50 procent, mens den kun er på 13 procent i matematik i 6. klasse. I afkodning (profilområde 2 i dansk læsning) ligger graden af gentestning generelt højt.

Der er generelt ingen statistisk signifikant korrelation mellem graden af genanvendelse af opgaver og tiden mellem de to frivillige test. I de fleste profilområder er der en svag tendens til, at graden af genanvendelse falder des længere tid, der er mellem elevens to test, men der er også enkelte profilområder, hvor tendensen er den modsatte. Det må umiddelbart forventes, at graden af genanvendelse falder, hvis elevens dygtighed ændres, da eleven så vil få opgaver fra andre områder i opgavebanken.

Sammenhængen mellem graden af genanvendelse af opgaver og elevens dygtighed i det andet forsøg i de frivillige test ses af tabel 14.

En positiv korrelation betyder, des dygtigere eleven er målt på logit skalaen des større er graden af genanvendelse af opgaver i andet forsøg i de frivillige nationale test.

Specielt i afkodning (profilområde 2) og i tekstforståelse (profilområde 3) i dansk læsning samt i alle tre profilområder i engelsk er der en høj positiv korrelation. Des dygtigere eleven er, des større sandsynlighed er der for at eleven får opgaver i andet forsøg, som eleven også har haft i første forsøg. Dette hænger sandsynligvis sammen med, at specielt disse profilområder mangler svære opgaver i opgavebanken.

Tabel 14 Sammenhængen mellem graden af genanvendelse af opgaver og elevdygtigheden målt på logit skalaen i forsøg 2. Spearman korrelationen

Test	Antal elever	Profilområde 1	Profilområde 2	Profilområde 3
Dansk/læsning 2. klasse	6.057	0,07*	0,43*	0,01
Dansk/læsning 4. klasse	4.421	-0,03*	0,57*	0,48*
Dansk/læsning 6. klasse	5.134	-0,09*	0,65*	0,61*
Dansk/læsning 8. klasse	2.558	0,05	0,52*	0,70*
Matematik 3. klasse	7.590	0,21*	0,17*	0,10*
Matematik 6. klasse	5.440	0,01	0,07*	-0,01*
Engelsk 7. klasse	2.424	0,47*	0,28*	0,37*
Fysik/kemi 8. klasse	1.169	-0,09*	-0,02	0,01
Biologi 8. klasse	761	-0,17*	0,07*	0,02
Geografi 8. klasse	1.028	-0,11*	-0,03	-0,12*

* Statistisk signifikant forskellig fra 0

Når der er taget højde for elevens dygtighed bestemt ved første forsøg, er der generelt ingen statistisk signifikant sammenhæng mellem elevens dygtighed i andet forsøg og graden af gentestning. *Det kan således ikke påvises, at gentestning øger sandsynligheden for et bedre testresultat.*

Derudover er der en svag statistisk signifikant negativ sammenhæng mellem graden af genanvendelse af opgaver i andet forsøg og ændringen i elevdygtigheden fra første til andet forsøg. Des større forskel i den beregnede elevdygtighed mellem de to frivillige forsøg, des mindre er sandsynligheden for at eleven får de samme opgaver begge gange. Samlet er korrelationen dog kun på $-0,02$, hvilket kan skyldes, at elevernes dygtighed ikke ændres markant på logit skalaen mellem de to frivillige testforsøg.

Eleverne opnår i gennemsnit næsten den samme score på de opgaver, de genanvender (tabel 15). Hvis scoren i andet forsøg generelt var større end i første forsøg, da kunne det indikere, at eleven havde lettere ved at svare korrekt i andet forsøg. Dette er altså ikke tilfældet.

Tabel 15 Forskel mellem scoren i forsøg 1 og forsøg 2 på de opgaver eleven møder begge gange

Test	Profil- område	Antal elever	25 % fraktil	Gennem- snit	75 % fraktil
Dansk/læsning 2. klasse	1	5.532	0	0	1
	2	5.499	0	1	2
	3	5.182	0	1	1
	Samlet	6.027	-1	2	3
Dansk/læsning 4. klasse	1	3.666	0	0	1
	2	4.086	0	1	1
	3	3.838	0	0	1
	Samlet	4.404	0	1	3
Dansk/læsning 6. klasse	1	4.118	0	0	1
	2	4.929	0	1	2
	3	4.511	0	1	1
	Samlet	5.130	0	2	3
Dansk/læsning 8. klasse	1	2.354	0	0	1
	2	2.515	0	1	2
	3	2.317	-1	1	2
	Samlet	2.558	0	2	3
Matematik 3. klasse	1	5.786	0	0	1
	2	6.171	0	0	1
	3	6.456	0	0	1
	Samlet	7.479	-1	1	2
Matematik 6. klasse	1	3.905	0	0	1
	2	4.061	0	0	1
	3	4.099	0	0	1
	Samlet	5.302	0	0	1
Engelsk 7. klasse	1	1.986	0	0	1
	2	1.878	0	0	1
	3	2.424	0	0	1
	Samlet	2.424	-1	1	2
Fysik/kemi 8. klasse	1	829	-1	0	1
	2	837	-1	0	1
	3	919	-1	0	1
	Samlet	1.144	-1	0	1
Biologi 8. klasse	1	575	-1	0	1
	2	592	-1	0	1
	3	625	-1	0	1
	Samlet	744	-1	0	1
Geografi 8. klasse	1	886	-1	0	1
	2	776	0	0	1
	3	842	-1	0	1
	Samlet	1.009	-1	0	1