



Bilagsnotat til: De nationale tests måleegenskaber

Baggrund

Der er ti obligatoriske test á 45 minutters varighed i løbet af elevernes skoletid. Disse er fordelt på seks forskellige fag og seks forskellige klassetrin. Fire af de ti test er i dansk, læsning på fire forskellige klassetrin og to af testene er i matematik.

I dette notat vil nogle af måleegenskaber ved de nationale test blive belyst. Det drejer sig om:

- Testenes validitet
- Testenes statistiske sikkerhed
- Testenes reliabilitet

Testenes validitet

I de nationale test testes alene i færdigheder, som det er muligt at afprøve i en it-baseret og selvrettende test. Hver test tester i tre faglige områder, de såkaldte profilområder.

For at få en indikation af om testene samlet set ser ud til at måle det samme som andre tilsvarende test og prøver, kan man se på sammenhængen mellem elevernes testresultat i de nationale test og deres efterfølgende præstation i de relevante dele af folkeskolens prøver i 9. klasse.

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.

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 (tabel 1).

Tabel 1 Elevernes testresultat i de obligatoriske nationale test i dansk, læsning sammenholdt med karakteren fra folkeskolens prøve i dansk læsning. Andel elever (pct.)

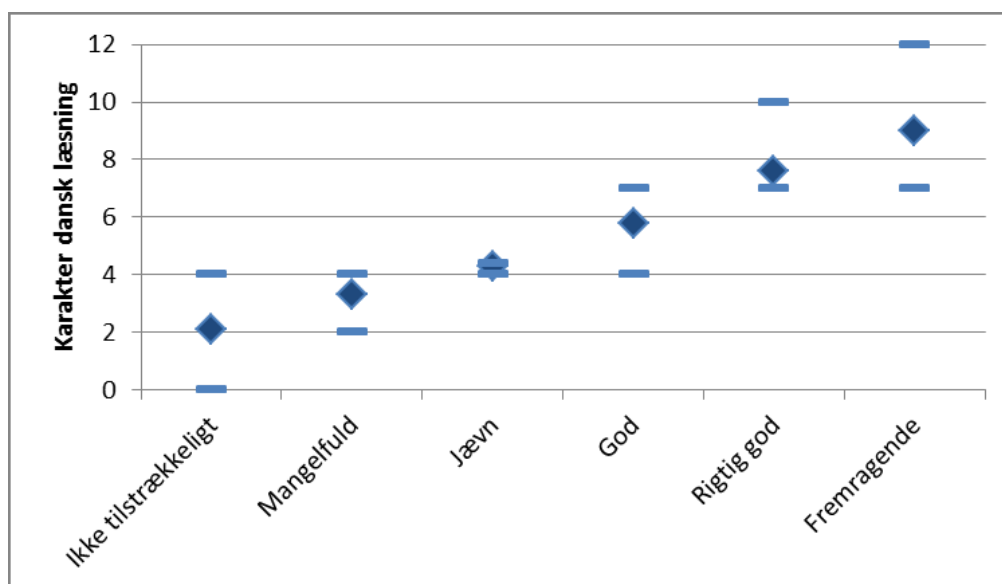
| Profil- område | Testresultat | Karakter | | | | | | | I alt |
|----------------------------------|--------------------|----------|----|----|----|----|----|----|-------|
| | | -3 | 0 | 2 | 4 | 7 | 10 | 12 | |
| Sprog forståelse 8. klasse | Ikke tilstrækkelig | 0 | 25 | 32 | 38 | 5 | 0 | 0 | 100 |
| | 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 8. klasse | 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 8. klasse | 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 8. klasse | 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 |

86 procent af de elever, der opnår en mangelfuld præstation i tekstforståelse, har fået karakteren 4 eller derunder i folkeskolens prøve i dansk læsning året efter, mens 13 procent har fået karakteren 7. 81 procent af de elever, der opnår en jævn præstation i tekstforståelse, har fået karakteren 4 eller 7 i folkeskolens prøve i dansk læsning året efter. 72 procent af de elever, der opnår en fremragende præstation i tekstforståelse, har mindst fået karakteren 10 året efter.

68 procent af de elever, der opnår en ikke tilstrækkelig præstation i den samlede vurdering i dansk læsning 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.

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 1. Gruppen af elever, der opnår en jævn præstation i samlet vurdering i de obligatoriske nationale test i dansk, læsning 8. klasse får i gennemsnit 4,3 ved folkeskolens prøve året efter i dansk læsning. Mens gruppen af elever, der opnår en rigtig god præstation i dansk, læsning 8. klasse i gennemsnit får 7,6 ved folkeskolens prøve året efter.

Figur 1 De nationale test i dansk læsning 8. klasse og folkeskolens prøve i 9. klasse



Anm.: Gennemsnitskarakter (firkant) samt 25 pct. og 75 pct. percentiler (vandret streg)

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.

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. Sammenhængen er ikke så stor som for dansk læsning, men i matematik er der også 3 år mellem de obligatoriske nationale test og folkeskolens prøve (tabel 2).

63 procent af de elever, der opnår en jævn præstation 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 opnår en fremragende præstation i tal og algebra, har fået karakteren 7 eller derover i folkeskolens prøve i matematik tre år efter.

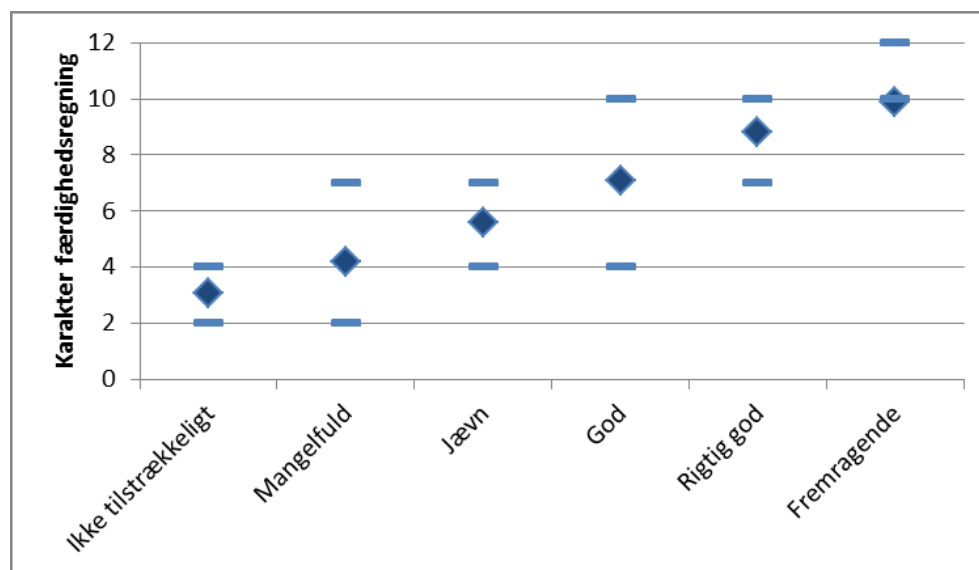
Tabel 2 Elevernes testresultat i de obligatoriske nationale test i matematik sammenholdt med karakteren fra folkeskolens prøve i matematiske færdigheder efter. Andel elever (pct.)

| Profil- område | Testresultat | Karakter | | | | | | | I alt |
|---|--------------------|----------|----|----|----|----|----|----|-------|
| | | -3 | 0 | 2 | 4 | 7 | 10 | 12 | |
| Tal og al- gebra 6. klasse | 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 og måling 6. klasse | 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 | 0 | 13 | 34 | 30 | 21 | 100 |
| | Fremragende | 0 | 0 | 12 | 9 | 18 | 35 | 26 | 100 |
| Matematik i anvendelse 6. klasse | 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 6. klasse | 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 |

55 procent af de elever, der opnår en ikke tilstrækkelig præstation i den samlede vurdering i matematik 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.

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 2. Gruppen af elever, der opnår en jævn præstation i samlet vurdering i de obligatoriske nationale test i matematik 6. klasse får i gennemsnit 5,6 ved folkeskolens prøve tre år efter i matematiske færdigheder. Mens gruppen af elever, der opnår en rigtig god præstation i samlet vurdering i matematik i 6. klasse i gennemsnit får 8,8 ved folkeskolens prøve i matematiske færdigheder tre år efter.

Figur 2 De nationale test i matematik 6. klasse og folkeskolens prøve i 9. klasse



Anm.: Gennemsnitskarakter (firkant) samt 25 pct. og 75 pct. percentiler (vandret streg)

Statistisk usikkerhed på testresultater fra de obligatoriske test 2016

Den statistiske usikkerhed (SEM) på et testresultat fra de obligatoriske nationale test i 2016 er i gennemsnit på 0,43. Den statistiske usikkerhed på den beregnede elevdygtighed er generelt størst for de dygtigste elever.

For hver elevs testresultat kan der beregnes et såkaldt sikkerhedsinterval. I forbindelse med beregning af sikkerhedsintervaller for gennemsnit, fx for en classes gennemsnit eller en skoles gennemsnit anvendes ofte et 95 procent sikkerhedsinterval. Et 95 procent sikkerhedsinterval er defineret som *gennemsnit ± 2*SE*.

Til angivelse af sikkerhedsinterval for et enkelt individs testresultat anvendes *testresultat ± 1*SEM*, hvilket svarer til et 67 procent sikkerhedsinterval.

Den beregnede elevdygtighed er den mest sandsynlige, men det kan ikke afvises med en vis sandsynlighed, at elevens dygtighed afviger en anelse fra dette.

Er en elevs testresultat således beregnet til at være en god præstation på den kriteriebaserede skala¹, da er 'God' det mest sandsynlige. Der er dog en

¹ Se appendiks for uddybning af de forskellige skalaer

sandsynlighed for, at elevens testresultat reelt ligger lige under ('Jævn') eller lige over ('Rigtig god').

Tilsvarende med testresultaterne præsenteret på den normbaserede percentilskala. Er en elevs score beregnet til fx 75, da er denne værdi den mest sandsynlige, men der er en sandsynlighed for, at elevens dygtighed er mindre eller større.

For hver elevs testresultat fra de obligatoriske nationale test i skoleåret 2015/2016 er der beregnet et sikkerhedsinterval på logit skalaen som *testresultat* $\pm 1*SEM$. Dette kan efterfølgende omregnes til et sikkerhedsinterval på såvel den kriteriebaserede skala som på percentilskalaen.

Testresultatet for 37 procent af de elever, der har opnået en god præstation i de nationale test, ligger med den beregnede statistiske sikkerhed² indenfor kategorien 'God', mens testresultatet for 27 procent af eleverne med den beregnede statistiske sikkerhed ligger indenfor kategorien 'God' eller 'Jævn', og testresultatet for 33 procent af eleverne ligger indenfor kategorien 'God' eller 'Rigtig god' (tabel 3).

43 procent af de elever, der har opnået en fremragende præstation, ligger med den beregnede statistiske sikkerhed indenfor kategorien 'Fremragende', mens testresultatet for 57 procent af eleverne ligger indenfor kategorien 'Fremragende' eller 'Rigtig god'.

² Der er anvendt et sikkerhedsinterval på $\pm 1*SEM$

Tabel 3 Sikkerhedsintervallet i forhold til elevernes dygtighed på den kriteriebaserede skala

| Kriterie | Andel elever | Ikke tilstrækkelig | Mangelfuld | Jævn | God | Rigtig god | Fremragende |
|--------------------|---------------------|--------------------|------------|-------|-------|------------|-------------|
| Ikke tilstrækkelig | 62 % | — x — | | | | | |
| | 38 % | — x — | | | | | |
| | 100 % ¹⁾ | | | | | | |
| Mangelfuld | 34 % | | — x — | | | | |
| | 47 % | — x — | | | | | |
| | 15 % | | — x — | | | | |
| | 3 % | — x — | | | | | |
| | 100 % ¹⁾ | | | | | | |
| Jævn | 5 % | | | — x — | | | |
| | 32 % | | — x — | | | | |
| | 26 % | | | — x — | | | |
| | 36 % | | — x — | | | | |
| | 100 % ¹⁾ | | | | | | |
| God | 37 % | | | | — x — | | |
| | 27 % | | | | — x — | | |
| | 33 % | | | | | — x — | |
| | 4 % | | | | — x — | | |
| | 100 % ¹⁾ | | | | | | |
| Rigtig god | 11 % | | | | | — x — | |
| | 27 % | | | | | — x — | |
| | 54 % | | | | | | — x — |
| | 8 % | | | | | — x — | |
| | 100 % ¹⁾ | | | | | | |
| Fremragende | 43 % | | | | | | — x — |
| | 57 % | | | | | | — x — |
| | 100 % ¹⁾ | | | | | | |

1) Enkelte testresultater (<0,02 pct.) er mere usikkert bestemt. Summen kan afvige fra 100 pct. pga. afrunding

Samlet gælder, at 28 procent af elevernes testresultater med den beregnede statistiske sikkerhed ligger indenfor det beregnede faglige niveau, mens 34 procent af elevernes testresultater ligger indenfor det beregnede niveau eller niveauet lige under og 29 procent af elevernes testresultater ligger indenfor det beregnede niveau eller niveauet lige over (tabel 4). Således ligger 91 procent af elevernes testresultater med den beregnede statistiske sikkerhed indenfor det beregnede faglige niveau eller i nabo kategorien. De resterende 9 procent af elevernes testresultater har en større usikkerhed, der betyder, at elevens faktiske niveau ikke kan afvises at ligge i både niveauet under og i niveauet over det målte niveau. Størstedelen af disse elever er elever, som vurderes til en jævn præstation.

Tabel 4 Den statistiske sikkerhed på elevens testresultat på den kriterie-baserede skala

| Elevens testresultat ligger med stor sandsynlighed i ... | | | |
|--|--|---|--|
| ... det beregnede faglige niveau | ... det beregnede faglige niveau eller niveauet lige under | ... det beregnede faglige niveau eller niveauet lige over | ... det beregnede faglige niveau eller niveauet lige over eller lige under |
| 28 pct. | 34 pct. | 29 pct. | 9 pct. |

Anm: Enkelte testresultater (<0,02 pct.) er mere usikkert bestemt

Sikkerhedsintervallet³ udtrykt på den normbaserede percentilskala er på ± 12 point. Dette svarer til, at en elevs testresultat med stor sandsynlighed ligger i intervallet testresultat $\pm 12^4$ (tabel 5). Det er vigtigt at understrege, at elevens beregnede score er den mest sandsynlige værdi, men der er en vis sandsynlighed for, at den reelle score afviger fra denne.

Tabel 5 Sikkerhedsintervallet i forhold til elevernes dygtighed på percentilskalaen

| Fag og klassetrin | Profilmråde | Samlet | Elevens dygtighed på percentilskalaen | | | | | | | | | |
|------------------------|-----------------------------|--------|---------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| | | | 1-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 90-100 |
| Læsning 2. klasse | Sprogforståelse | 14 | 3 | 9 | 13 | 16 | 18 | 18 | 17 | 15 | 11 | 6 |
| | Afkodning | 8 | 2 | 4 | 6 | 8 | 9 | 10 | 10 | 9 | 6 | 3 |
| | Tekstforståelse | 8 | 2 | 5 | 6 | 8 | 10 | 11 | 12 | 10 | 7 | 4 |
| Læsning 4. klasse | Sprogforståelse | 13 | 3 | 9 | 13 | 16 | 18 | 19 | 18 | 16 | 11 | 5 |
| | Afkodning | 10 | 3 | 7 | 10 | 12 | 14 | 15 | 14 | 12 | 8 | 4 |
| | Tekstforståelse | 10 | 3 | 6 | 9 | 12 | 14 | 15 | 15 | 13 | 9 | 4 |
| Læsning 6. klasse | Sprogforståelse | 14 | 4 | 10 | 14 | 17 | 19 | 19 | 19 | 16 | 12 | 5 |
| | Afkodning | 11 | 3 | 8 | 11 | 13 | 15 | 16 | 16 | 14 | 10 | 5 |
| | Tekstforståelse | 11 | 3 | 7 | 11 | 15 | 17 | 17 | 16 | 14 | 10 | 4 |
| Læsning 8. klasse | Sprogforståelse | 16 | 4 | 11 | 16 | 20 | 22 | 23 | 22 | 20 | 15 | 4 |
| | Afkodning | 11 | 3 | 9 | 12 | 15 | 15 | 15 | 13 | 11 | 8 | 4 |
| | Tekstforståelse | 11 | 2 | 7 | 11 | 14 | 16 | 17 | 17 | 14 | 10 | 4 |
| Matematik 3. klasse | Tal og algebra | 11 | 3 | 7 | 11 | 13 | 14 | 15 | 15 | 13 | 9 | 3 |
| | Geometri | 13 | 4 | 10 | 15 | 18 | 19 | 19 | 17 | 14 | 10 | 4 |
| | Stat. og sands ¹ | 11 | 3 | 8 | 13 | 16 | 17 | 16 | 14 | 11 | 7 | 3 |
| Matematik 6. klasse | Tal og algebra | 13 | 4 | 11 | 16 | 19 | 20 | 20 | 17 | 15 | 11 | 4 |
| | Geometri | 13 | 4 | 9 | 13 | 16 | 18 | 18 | 18 | 16 | 11 | 4 |
| | Mat. i anvend. ² | 11 | 3 | 9 | 13 | 15 | 16 | 17 | 16 | 13 | 9 | 4 |

1) Statistik og sandsynlighed 2) Matematik i anvendelse

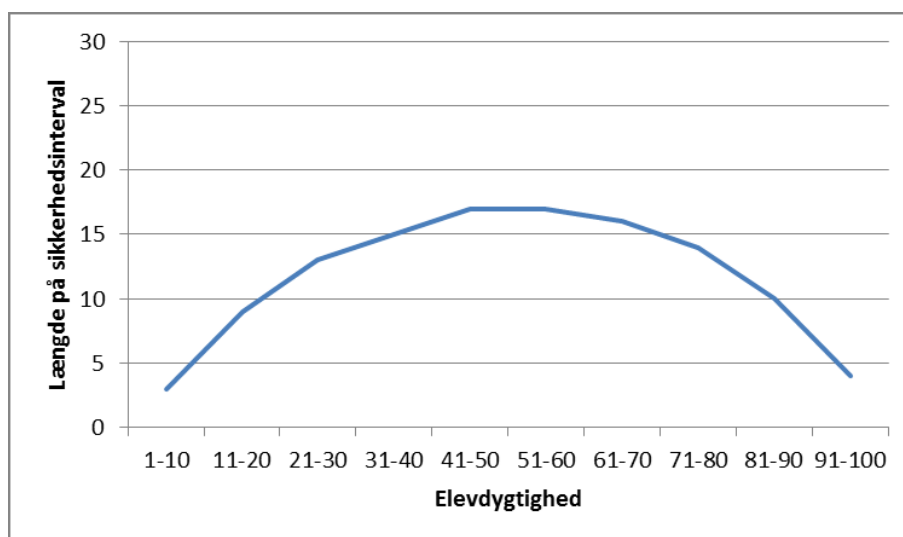
³ Der er anvendt et sikkerhedsinterval på $\pm 1 \cdot \text{SEM}$

⁴ Der er tale om en lille tilnærmelse, da sikkerhedsintervallerne på percentilskalaen ikke er helt symmetriske. Endvidere er sikkerhedsintervallerne afgrænset til intervallet 1-100.

Den statistiske sikkerhed udtrykt på percentilskalaen afhænger ikke kun af fag, klasstrin og profilområde men også af elevens testresultat. Fx er sikkerhedsintervallet for en elev, der har opnået 35 i tekstforståelse i 4. klasse, lig med ± 12 , dvs sikkerhedsintervallet for denne elevs testresultat kunne være [23 - 47], mens sikkerhedsintervallet for en elev, der har opnået 85 i tekstforståelse i 4. klasse, er lig med ± 9 , dvs sikkerhedsintervallet for denne elevs resultat kunne være [76 - 94].

Længden på sikkerhedsintervallet udtrykt på percentilskalaen er størst i midten af den normbaserede skala fra 30-70, mens længden på sikkerhedsintervallet er mindst for de dygtigste og de mindre dygtige elever (figur 3).

Figur 3 Den statistiske sikkerhed på elevens testresultat på percentilskalaen



Den statistiske usikkerhed på testresultatet på logit skalaen⁵ er størst blandt de dygtigste elever. Men når resultaterne præsenteres på percentilskalaen sløres denne egenskab. Længden på sikkerhedsintervallet udtrykt på percentilskalaen afspejler således ikke udelukkende den statistiske usikkerhed på selve testresultatet men er i højere grad et udtryk for percentilskalaens egenskaber.

Testenes reliabilitet

Reliabiliteten er et udtryk for, i hvor høj grad en elev opnår det samme testresultat, hvis eleven tager den samme test to gange. Gentagelsen skal ske uden, at eleven kan huske det første testforløb og uden, at eleven har lært af den første test eller lært nyt mellem de to testafviklinger. Dette er vanskel-

⁵ Se appendiks for uddybning af de forskellige skalaer

ligt i pædagogiske test, hvorfor der primært anvendes simuleringer til at belyse graden af reliabiliteten.

For hvert profilområde er der simuleret to gentagne elevforløb for 5.000 elever⁶. Hvert testforløb består af 20 opgaver i hvert profilområde.

Simuleringerne har den fordel, at de ikke er påvirket af elevernes aktuelle testadfærd. Simuleringer er derfor mere velegnede til at afprøve et test- og prøvesystem som det, der er anvendes i de nationale test.

I alle profilområder er der en statistisk signifikant positiv sammenhæng mellem elevdygtigheden bestemt ved to simulerede testforløb. Udtrykt på percentilskalaen er forskellen i den beregnede elevdygtighed mellem de to simuleringer i gennemsnit lig nul med et interkvartil⁷ range på [-8; +8].

Sammenhængen mellem elevernes testresultater i de to gentagelser er udtrykt ved korrelationskoefficienten⁸ (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 Korrelationen mellem elevdygtigheden bestemt ved to simulerede testforløb

| Test | Profilområde 1 | Profilområde 2 | Profilområde 3 |
|-------------------------|----------------|----------------|----------------|
| Dansk læsning 2. klasse | 0,78 | 0,93 | 0,91 |
| Dansk læsning 4. klasse | 0,82 | 0,89 | 0,90 |
| Dansk læsning 6. klasse | 0,82 | 0,86 | 0,87 |
| Dansk læsning 8. klasse | 0,84 | 0,87 | 0,88 |
| Matematik 3. klasse | 0,90 | 0,86 | 0,82 |
| Matematik 6. klasse | 0,89 | 0,86 | 0,89 |

Med undtagelse af de nationale test i sprogforståelse (profilområde 1) i dansk læsning 2. klasse ligger alle korrelationerne i intervallet 0,82 - 0,93⁹. Medianen er på 0,87.

⁶ For hvert profilområde er der simuleret 5.000 elevers testforløb. Eleverne har fra start en dygtighed angivet på logit skalaen. De 5.000 elevers dygtighed følger en Normalfordeling med middelværdi og spredning som observeret ved de obligatoriske test. Eleverne er således en repræsentativ afspejling af en faktisk elev population

⁷ 25 pct. og 75 pct. percentiler

⁸ Korrelationerne er beregnet på baggrund af testresultater på logit skalaen

⁹ Guideline til vurdering af korrelations koefficienter: '0,0-0,2'=meget svag; '0,2-0,4'=svag; '0,4-0,6'=moderat; '0,6-0,8'=stærk; '0,8-1,0'=meget stærk. (Evans, J.D. 1996: Straightforward statistics for the behavioral sciences)

De nationale test er som udgangspunkt udviklet som et redskab til de obligatoriske målinger på bestemte klassetrin. Muligheden for at gennemføre frivillige nationale tests i efterårssemesteret har dog gennem de seneste år været stigende. I alt gennemførte knap 320.000 elever i efteråret 2015 de frivillige nationale test. Ca. 35.000 af dem gennemførte to på hinanden følgende frivillige tests i samme fag.

Når man gennemfører to på hinanden følgende tests med få ugers mellemrum¹⁰ er der mange faktorer, der kan spille ind i forhold til, om man kan sammenligne de to testresultater. Lærerens instruktioner og formålet med de to hurtige testafviklinger, elevens motivation og koncentration samt stabiliteten af lokalt it-udstyr er nogle af de forhold, der kan påvirke et testforløb. Det er derfor ikke usædvanligt, at observerede test-retest korrelationer er lavere end korrelationer baseret på simulerede data. Gentagne test er således vanskelige, at anvende som et direkte mål for reliabiliteten af de pågældende test.

Som supplement til de simulerede elevforløb er der dog også set på korrelationen mellem testresultaterne for de elever, der tog den samme test indenfor én uge. Korrelationen mellem elevers testresultater fra to gentagne testforløb er lavere (tabel 7) end korrelationerne beregnet ved simuleringerne.

Tabel 7 Korrelationen mellem elevdygtigheden bestemt ved to forsøg i de frivillige test

| Test | Profil- område 1 | Profil- område 2 | Profil- område 3 |
|-------------------------|---------------------|---------------------|---------------------|
| Dansk læsning 2. klasse | 0,64 | 0,83 | 0,79 |
| Dansk læsning 4. klasse | 0,73 | 0,85 | 0,77 |
| Dansk læsning 6. klasse | 0,60 | 0,81 | 0,74 |
| Dansk læsning 8. klasse | 0,66 | 0,85 | 0,72 |
| Matematik 3. klasse | 0,75 | 0,67 | 0,78 |
| Matematik 6. klasse | 0,63 | 0,65 | 0,68 |

Anm.: Testresultater fra de elever der har gennemført den samme test to gange i efteråret 2015 med højst én uges mellemrum

¹⁰ I gennemsnit var der 20 dage mellem

Appendiks

Anvendte skalaer til præsentation af elevens beregnede dygtighed

Elevdygtigheden og den tilhørende usikkerhed beregnes i den adaptive algoritme i test- og prøvesystemet på en **logit skala**, som i princippet går fra minus til plus uendelig, men som i praksis går fra minus 7 til plus 7. I de fleste profilområder er intervallet dog snævrere.

Siden skoleåret 2009/2010 har resultaterne fra de nationale test været formidlet på den normbaserede skala (1-100), der er en **percentilskala**. Percentilskalaen er dannet på baggrund af fordelingen af elevernes testresultater i 2010, hvor en percentilværdi på fx 40 svarer til den elevdygtighed på logit skalaen, hvor 40 procent af testresultaterne i 2010 lå under. En percentilværdi på 90 svarer til den elevdygtighed, hvor 90 procent af eleverne i 2010 scorede under, dvs kun 10 procent af eleverne i 2010 opnåede et bedre testresultat.

Logit skalaen har den fordel, at det er en såkaldt ækvidistant skala. Forskellen mellem to værdier har samme betydning overalt på skalaen. Således er fx en forbedring eller en forskel fra -2 til -1 den samme som en forbedring fra 0 til +1.

Dette gælder ikke for percentilskalaen. Her vil en forskel eller en forbedring på fx 10 point svarer til en relativ beskeden ændring i elevdygtigheden på logit skalaen, hvis der er tale om en middeldygtig elev (30-70), mens det svarer til en større ændring i elevdygtigheden på logit skalaen, hvis der er tale om en meget dygtig elev (80-100) eller en mindre dygtig elev (1-20).

Siden skoleåret 2014/2015 har resultaterne i dansk læsning og matematik desuden været præsenteret på den **kriteriebaserede skala**. Ved den kriteriebaserede scoring omregnes elevens dygtighed på logit skalaen til en forventet score på et antal udvalgte opgaver. Opgavekommissionerne har fastsat kriterier for, hvilket fagligt niveau en score stemmer overens med. I hvert profilområde er defineret seks faglige niveauer: ikke tilstrækkelig præstation, mangelfuld præstation, jævn præstation, god præstation, rigtig god præstation og fremragende præstation.