



Faktaark om de nationale test

De nationale test – klassetrin, fag og profilområder

- Der er ti obligatoriske test i folkeskolen, hvor hver test består af tre faglige profilområder. Fx består testen i dansk læsning af profilområderne sprogforståelse, afkodning og tekstforståelse.

Fag og klassetrin	1.	2.	3.	4.	5.	6.	7.	8.	9.
Dansk, læsning									
Matematik									
Engelsk									
Fysik/kemi									
Biologi									
Geografi									
Dansk som andetsprog									

- Obligatorisk test målrettet klassetrinnet
- Frivillig test målrettet klassetrinnet
- Frivillig test målrettet klassetrinnet over eller under

De nationale test er it-baserede, selvscorende og adaptive

- At testene er adaptive betyder, at opgaverne i et testforløb udvælges, så de bedst muligt passer til elevens dygtighed undervejs i forløbet. Hvis eleven svarer rigtigt på en opgave, får eleven næste gang en lidt sværere opgave. Svarer eleven forkert på en opgave, får eleven næste gang en lidt lettere opgave.
- Når testen er afsluttet, beregnes elevens dygtighed i hvert af profilområderne.
- Dygtigheden beregnes ved anvendelse af en statistisk model, Rasch modellen, hvor opgavernes sværhedsgrad indgår sammen med elevens besvarelse af opgaverne.

Usikkerhed på den beregnede elevdygtighed

- Der er statistisk usikkerhed på alle test og prøver. Resultater fra de nationale test er også forbundet med statistisk usikkerhed.
- Generelt gælder det, at des flere opgaver en test eller prøve indeholder, des mindre er den statistiske usikkerhed. Dette gælder også i de nationale test. Des flere opgaver, eleven når at besvare, des mindre usikkerhed er der på den beregnede dygtighed.
- Der er en sammenhæng mellem antal opgaver og den tid, en test tager.
- Usikkerheden kan primært reduceres ved at forlænge testtiden og lade eleven svare på flere opgaver.
- I adaptive test, hvor opgavernes sværhedsgrad passer til elevernes dygtighed, er den statistiske usikkerhed mindre end i almindeligt lineære test med samme antal opgaver. I lineære test får alle elever de samme opgaver uanset elevernes dygtighed.

Visning for læreren af, hvor sikkert en elevs dygtighed er

- I testsystemet vises den statistiske usikkerhed sammen med elevens testresultat.
- Læreren kan se, hvor mange opgaver den enkelte elev har besvaret, hvad eleven har svaret på hver opgave, den tid eleven har brugt på opgaverne samt den beregnede usikkerhed.

Betydning af SEM og grænseværdi for statistisk usikkerhed

- I testsystemet anvendes den statistiske usikkerhed, SEM (Standard Error of Measurement), til at angive om en elevs dygtighed er målt sikkert nok til at afslutte testen.
- Der anvendes en grænseværdi af SEM på 0,55. Denne grænse har været fastsat fra opstarten i 2010 og sikrer, at de fleste elever kan færdiggøre testen på 45 minutter. Jo lavere værdi, jo sikrere et det, at elevens dygtighed er målt korrekt.
- Når den statistiske usikkerhed på elevens resultat kommer under 0,55 eller eleven har besvaret 29 opgaver i et profilområde, er testsystemet indrettet til ikke at stille flere opgaver i dette profilområde, før usikkerheden på elevens resultat er under 0,55 i alle tre profilområder (eller eleven har svaret på 29 opgaver i hvert af profilområderne). Herefter får læreren besked om, at testen kan afsluttes. Testen kan dog tidligst afsluttes, når de afsatte 45 minutter er gået.
- Læreren kan vælge at forlænge en elevs test ud over de afsatte 45 minutter for at øge sikkerheden, men det er op til læreren at afgøre i det enkelte tilfælde. I knap 7 procent af tilfældene er den statistiske usikkerhed på elevens resultat større end 0,55 selv efter forlængelse.

Forskelle i usikkerhed – de dygtigste elever

- Den statistiske usikkerhed er størst for de dygtigste elever.
- I den bagved liggende statistiske model (Rasch modellen) gælder, at den statistiske usikkerhed blandt andet afhænger af elevens dygtighed. Denne egenskab betyder, at den statistiske usikkerhed ofte vil være størst for de dygtigste og for de svageste elever. Usikkerheden er mindst omkring gennemsnittet af elevdygtighederne.

- Derudover kan det skyldes et utilstrækkeligt antal svære opgaver til de dygtigste elever. Den statistiske usikkerhed bliver mindst, når opgavernes sværhedsgrad passer til elevernes dygtighed. Når der ikke er tilstrækkeligt mange svære opgaver, er det mere vanskeligt at bestemme de dygtigste elevers præcise dygtighedsniveau.
- Endelig kan det skyldes, at de svære opgaver tager længere tid at løse, hvorfor eleverne ikke kan nå at besvare så mange af disse opgaver.

Den statistiske usikkerhed er forskellig fra fag til fag

- Den observerede statistiske usikkerhed på elevens testresultat er forskellig fra fag til fag.
- Dette skyldes, at der er profilområder, hvor eleverne når at besvare flere opgaver end andre. Dette gælder blandt andet profilområder, hvor der er flere opgaver med flere delopgaver.
- En opgave med flere delopgaver (polytom opgave) tager generelt kortere tid for eleven at besvare end et tilsvarende antal enkelt-opgaver. Derfor kan eleven nå at besvare flere spørgsmål i disse profilområder, og den statistiske usikkerhed bliver mindre.

Statistisk usikkerhed i de nationale test sammenlignet end andre test og prøver

- Der er statistisk usikkerhed på alle test og prøver. Resultater fra de nationale test er også forbundet med statistisk usikkerhed.
- I de nationale test når eleverne i gennemsnit at besvare 23 delopgaver inden for hvert af de tre profilområder. Inden for rammerne af en test på 23 opgaver er der et begrænset forbedringspotentiale med hensyn til reduktion af den statistiske usikkerhed.
- Mange test og prøver indeholder flere opgaver, og der er afsat længere tid. Derfor må det forventes, at sikkerheden i elevernes testresultat er større i disse. Til sammenligning indeholder folkeskolens digitale selvrettende prøver i 9. klasse ca. 50 delopgaver.

Hvad betyder validitet?

- Validitet dækker over flere ting.
- Intern validitet dækker over, om en test måler det, den skal måle.
- Ekstern validitet dækker over, om testens resultater kan generaliseres til andre tilsvarende resultater (fx prøver). Ekstern validitet er undersøgt ved at sammenholde elevers resultater fra de nationale test med tilsvarende resultater fra folkeskolens prøver.
- Evalueringen viser en høj ekstern validitet.

Hvad betyder reliabilitet?

- Reliabilitet er et udtryk for i hvor høj grad testens rangordning af eleverne fra de mindst dygtige til de dygtigste er troværdig.
- Reliabiliteten afhænger både af testens statistiske sikkerhed og af spredningen mellem elevernes dygtighed. Jo større spredning der er mellem elevernes dygtigheder, jo nemmere er det at adskille elevernes faglige dygtighed ud fra en test.
- Tilsvarende gælder det, at hvis usikkerheden på den enkelte elevs beregnede dygtighed er lille, da vil reliabiliteten også være høj.

Fastsættelse af opgavernes sværhedsgrad

- I de nationale test afprøves alle opgaver af cirka 700 elever, inden de anvendes.
- Opgaveafprøvningerne foregår hvert år blandt tilfældigt udvalgte skoler.
- Opgaveafprøvningen foregår som en lineær test, hvor de 700 elever får de samme opgaver – uanset elevernes dygtighed. Opgaverne afprøves af elever på det klassetrin, testen er målrettet.
- Efter opgaveafprøvningen analyserer Styrelsen for It og Læring elevernes besvarelser, og opgaver, der ikke passer til den statistiske model, kasseres. Ligeledes kasseres opgaver, hvor opgavens sværhedsgrad afhænger af elevens køn o.a.
- Opgaver, der godkendes, får deres opgavesværhedsgrad beregnet på baggrund af de indsamlede besvarelser fra afprøvningen.