

Opgavebeskrivelse for evaluering af de nationale test

Baggrund

Folketinget vedtog i marts 2006 indførelsen af de nationale test. Den første obligatoriske testrunde blev gennemført i foråret 2010. De nationale test var ét blandt flere elementer i et lovforslag fra december 2005 om fornyelse af folkeskolen for at forbedre det faglige niveau blandt eleverne.

De nationale test tjener to formål: Med vedtagelsen af L101 indførtes de nationale test som et pædagogisk redskab til lærerne. Med vedtagelsen af L170 blev de nationale test desuden givet et styringsformål.

De nationale test har følgende grundlæggende karakteristika:

- Hver test består af tre faglige profilområder.
- De er it-baserede.
- De er adaptive, hvilket vil sige, at de tilpasser sig elevens niveau undervejs i testforløbet.
- De er selvscorende.
- Der gives en tilbagemelding pr. profilområde samt en samlet vurdering.
- En test kan gennemføres på én lektion (45 min.).

De nationale test blev senest evalueret i 2013. Folketinget blev som opfølgning på evalueringen i 2013 lovet, at der skulle igangsættes en ny evaluering efter en femårig periode. Evalueringen skal danne grundlag for en ny redegørelse til folkettinget og for at vurdere behovet for eventuelle ændringer. Der henvises til bilag 1 for yderligere baggrund om de nationale test.

Opgavens formål

Formålet med evalueringen af de nationale test er at få viden om, hvorvidt de nationale tests indhold og udformning har styrket skolernes evalueringskultur og derigennem elevernes faglige niveau.

Evalueringen skal belyse styrker såvel som svagheder omkring indholdet og brugen af de nationale test i folkeskolen.

Evalueringen skal give et vidensgrundlag, der kan danne grundlag for redegørelse til folkettinget og danne afsæt for det fremadrettede arbejde med udvikling og brug af de nationale test i folkeskolen.

Opgavens indhold og metode

De nationale test tjener flere formål og skal med sine resultater give mening for mange forskellige brugere på flere forskellige niveauer. Det er derfor afgørende for brugbarheden af evalueringen, at den gives et summativt såvel som formativt sigte, og at den afvejer de heterogene forhold med de respektive overordnede formål for testene rettet mod mange aktører med henblik på, at der kan træffes et afvejet og kvalificeret valg om den fremadrettede brug og udvikling af de nationale test.

Opgaven består overordnet af to dele:

- 1) Validering af teknisk beregning
- 2) Undersøgelse af betydningen og brugen af de nationale test

Delopgave 1: En analyse af den statistiske usikkerhed, reliabiliteten og øvrige måleegenskaber forbundet med de nationale test.

STIL gennemfører en analyse af den statistiske usikkerhed på de beregnede elevdygtigheder og af reliabiliteten af testene.

Konkret bliver følgende evalueringsspørgsmål besvaret:

1. *Regner de nationale test rigtigt?*

Ved besvarelse af spørgsmålet skal den kritik der rejses af den statistiske sikkerhed og reliabiliteten i de nationale test adresseres. Herunder skal det klarlægges om:

- a. opgavernes sværhedsgrader stadig er korrekte?
- b. opgaverne fortsat passer til Rasch-modellen?
- c. det er muligt at forbedre den adaptive algoritme med henblik på at reducere den statistiske usikkerhed?

2. Det skal afdækkes, om sikkerheden i målingerne af elevernes færdigheder forbedres ved at kombinere resultater fra forskellige profilområder? Herunder sigter spørgsmålet på at klarlægge følgende:

- a. Kan det påvises, at profilområderne måler forskellige aspekter af den samme bagvedliggende færdighed?
- b. Som følge af spørgsmål a: Kan testresultaterne fra profilområderne slås sammen og dermed forbedre sikkerheden i testene?

Delopgavens metode: Spørgsmålene et og to løses internt i Undervisningsministeriet ved Styrelsen for IT og Læring (STIL). Undersøgelsen heraf fordrer ikke indsamling af nye data, men skal gennemføres ved analyse af eksisterende data samt ved at udarbejde yderligere dokumentation for testenes egenskaber.

STIL udarbejder dokumentation for den anvendte metode til bestemmelse af opgavernes sværhedsgrader herunder de enkelte faser i den statistiske analyse. Der foretages analyse af, hvorvidt opgavernes sværhedsgrader er ændret over årene, og i hvilket omfang der tages højde for dette i den adaptive algoritme.

Endvidere vil mulige forbedringstiltag af den adaptive algoritme blive beskrevet herunder effekten på den statistiske usikkerhed.

Muligheden for at samle testen i de tre adskilte profilområder til ét samlet profilområde undersøges ved analyse af eksisterende elevbesvarelser.

STIL involverer eksterne forskere i analyserne. Evaluator opgave i forbindelse med delopgave 1 er at validere resultaterne af STILs dokumentation og analyser ved brug af nationale og nordiske forskere på området. STUK skal godkende reviewgruppen.

Delopgave 2: Undersøgelse af brugen af de nationale test

Der skal gennemføres en kortlægning og analyse af brugen af de nationale test blandt elever, lærere, ledere, kommunalt forvaltnings- og politisk niveau og på statsligt forvaltnings- og politisk niveau. Konkret skal følgende evalueringsspørgsmål besvares:

1. *Findes der test i Danmark eller andre lande, som med samme tidsforbrug måler mere præcist end de nationale test?*

Der skal udarbejdes en kortlægning og komparativ analyse af det danske og internationale testlandskab, der omfatter

- a. En analyse af målepræcisionen i andre danske test i forhold til hvad der er beskrevet i rammedokumenter om skolens formål og målsætninger og i forhold til tidsforbruget ved de analyserede test?
- b. En kortlægning af andre landes nationale testredskaber og en analyse af præcisionen i disse sammenlignet med de nationale test og set i forhold til tidsforbruget med testredskaberne.
- c. En kortlægning af internationale erfaringer med raschscorede lineære test og erfaringer med en kombination af adaptive og lineære test

2. *I hvilket omfang er der sammenhæng mellem på den ene side skolens formål og målsætninger beskrevet i rammedokumenter og på den anden side indholdet af de nationale test?*

Der skal udarbejdes en kortlægning og analyse af de sammenhænge, der er mellem folkeskolens formål og indholdet af de nationale test, der omfatter

- a. En analyse af om de nationale test måler centrale dele af faget og fagenes formål jf. Fælles Mål. Der tages udgangspunkt i fagenes formål, og de centrale dele af fagene udvælges i samarbejde med STUK.
- b. En gennemgang af det faglige indhold der testes på i de nationale test samt en empirisk analyse af samvariationen i elevers præstationer i testene og andre aspekter af fagets formål.

3. *Oplever lærerne, at den information som de nationale tests producerer, er nyttig og handlingsvejledende i det pædagogiske arbejde?*

Herunder ønskes som minimum en analyse af anvendeligheden for lærere og ledere af vejledningerne knyttet til de nationale test, der omfatter

- a. En analyse af den pædagogiske og didaktiske informationsværdi af oplysninger knyttet til de nuværende profilområder for lærere og ledere.
- b. En kortlægning og analyse af i hvilket omfang de kriteriebaserede og normbaserede resultater giver oplysninger, som lærere, vejledere og ledere kan bruge til at vurdere, hvor eleven ligger i forhold til skolens formål og målsætning?
- c. En undersøgelse af hvilke tilbagemeldinger, der er tilgængelige for lærere, herunder hvilken type tilbagemeldinger lærerne foretrækker.

4. *Hvordan forstår lærerne, hvad resultaterne af de nationale test viser?*

En kortlægning og analyse af lærernes forståelse og fortolkning af de nationale testresultater, der omfatter

- a. En kortlægning af i hvilken grad lærerne har kendskab til vejledningsmaterialet om de nationale test
 - b. En kortlægning af, hvorvidt de oplever, at den information som de nationale test producerer, er nyttig og handlingsanvisende i det pædagogiske arbejde.
 - c. En analyse af i hvilken grad vejledningerne svarer til det som de nationale tests viser.
 - d. En analyse af den sammenhæng der er mellem a, b og c.
5. *Når de nationale test fokuserer på delaspekter af specifikke fagområder, hvordan præger valget af disse delaspekter så den måde, som det enkelte fag opfattes, realiseres og prioriteres?*
En analyse af hvordan undervisningen afspejler profilområderne i de enkelte test, der som minimum omfatter:
- a. En analyse af om og i givet fald hvordan lærerne oplever, at de nationale test ved deres blotte eksistens, indhold og form, påvirker lærernes undervisning; dels i klassesammenhæng, dels i forhold til den enkelte elev.
6. *Hvordan er evalueringspraksis på skolerne?*
En kortlægning og analyse af den eksisterende evalueringspraksis på skolerne, der som minimum omfatter:
- a. En kortlægning af hvilke redskaber lærere og ledere bruger til at få indblik i elevernes faglige niveau.
 - b. En kortlægning af hvordan dialoger omkring testresultaterne foregår.
 - c. En kortlægning af hvordan testresultaterne indgår i den pædagogiske praksis i forhold til eleverne.
 - d. En kortlægning og analyse af hvordan lærernes oplevelse af sammenhængen mellem de nationale test på den ene side og på den anden side det pædagogiske arbejde og skolens evalueringspraksis i øvrigt.
 - e. En kortlægning af, hvad forvaltningsniveau, ledere, vejledere og læreres oplevelse af udviklingen af evalueringspraksis er over de seneste år
7. *I hvilket omfang og hvordan arbejder skoleledere og lærere sammen om at tolke testresultater og bruge dem i forhold til udvikling af den pædagogiske praksis?*
Der udarbejdes en kortlægning af læreres og lederes samarbejde om de nationale test og testresultater og tolkningen deraf. Kortlægningen skal som minimum omfatte:
- a. En kortlægning og analyse af hvordan vejledere og konsulenter inddrages i samarbejdet, og hvem der inddrager disse?
 - b. En analyse af hvordan samarbejdet mellem ledere og lærere om testene påvirker lærernes didaktik
8. *Hvordan anvender og oplever skoleledelsen de nationale test i ledelsesarbejdet og i skolernes evalueringspraksis i øvrigt?*
Der skal udarbejdes en kortlægning og analyse af ledelsens anvendelse af de nationale test i relation til forvaltningslaget og pædagogiske og strategiske ledelse af skolen, der som minimum omfatter:
- a. En kortlægning af samarbejdet mellem skoleledelsen og forvaltningen på baggrund af og omkring testresultaterne.

- b. En kortlægning og analyse af hvilken rolle de nationale testresultater spiller for ledelsen af den enkelte lærer, og hvordan lærerne tilsvarende oplever, at ledelsen anvender testresultaterne?
- c. En kortlægning af om de nationale test spiller en anden rolle end eventuelle øvrige test på skolen i ledelsesarbejdet og i givet fald hvilken.

9. *Hvad er elevernes oplevelse af, hvordan lærerne forklarer om de nationale test, herunder testsituationen, de nationale tests formål og anvendelse?*

Der skal gennemføres en analyse af situationen, der omkranser testsituationen således der analyseres på sammenhænge mellem lærernes rammesætning af testene og elevernes testadfærd. Analysen skal gennemføres i indskoling på mellemtrin og i udskoling og skal som minimum indeholde

- a. En kortlægning af elevernes oplevelse af lærernes forklaring af de nationale test
- b. En analyse af den introduktion og rammesætning lærerne giver før gennemførelse af de nationale test
- c. En analyse af lærernes adfærd i forbindelse med gennemførelse af de nationale test
- d. En analyse af sammenhængen til elevernes oplevelse af testsituationen både før, under og efter gennemførelsen af testene.
- e. En analyse af elevernes testadfærd i forbindelse med afviklingen af testene og sammenhænge til lærernes rammesætning og adfærd i testsituationen.

10. *Hvad er virkningen af de nationale test på elevernes faglige niveau?*

Der skal gennemføres en analyse af hvilken betydning de nationale test har og har haft for udviklingen i elevernes faglige niveau. Analysen skal i videst muligt omfang isolere betydningen af de nationale test fra andre indsatser med betydning for elevernes faglige resultater. Analysen skal som minimum

- a. Forholde sig til hvordan de nationale test som del af skolernes evalueringspraksis har og har haft betydning for elevernes faglige udvikling.
- b. Forholde sig til hvordan de nationale test har og har haft betydning for mål for elevernes læring, mål for lærernes og skolernes pædagogik og didaktik
- c. Forholde sig til hvordan de nationale test har og har haft betydning for lærerens tilrettelæggelse af undervisningen af den enkelte elev og undervisning af klassen.

11. *Får eleverne feedback efter test og i så fald, kan de bruge den i forhold til deres egen læring? Indgår eleverne i samarbejde om opfølgning på de nationale test?*

Der skal gennemføres en kortlægning af den feedback eleverne får efter testenes afholdelse, der som minimum skal indeholde

- a. En kortlægning af omfang og indhold af den feedback lærere giver til eleverne.
- b. En kortlægning af i hvilket omfang testresultaterne genbesøges og formativt danner grundlag for elevernes faglige udvikling.

12. *Hvordan vurderer forældre (som vidner til elevernes reaktioner og lærernes information) fordele og ulemper ved de nationale test?*

Der skal gennemføres en kortlægning af forældrenes vurdering af de fordele og ulemper der ser ved brugen af de nationale test. Kortlægningen skal som minimum omfatte

- a. Den opfattelse de har af nytten af de nationale test

- b. Den tilbagemelding og italesættelse de oplever fra elever og lærere
13. *Hvordan anvendes de nationale test institutionelt til styring?*
Der skal gennemføres en kortlægning af omfang og måden hvorpå skolebestyrelsen i samarbejde med skolens ledelse ser og anvender de nationale test til at sætte pædagogiske og strategisk retning for skolen. Denne skal som minimum omfatte
- Hvordan og i hvilket omfang anvender skolebestyrelserne resultaterne af de nationale test i arbejdet.
 - Hvilke andre data de nationale test kombineres med som grundlag for beslutninger om skolen strategisk og pædagogisk.
En kortlægning af, hvorvidt ledere og skolebestyrelser anvender muligheden for at følge faglig progression gennem hele skoleforløbet.
14. *Hvordan anvendes de nationale test på kommunalt styringsniveau?*
Der skal gennemføres en kortlægning og analyse af det kommunale niveaus anvendelse af testdata fra de nationale test i relation til styring af det samlede kommunale skolevæsen og i relation til den enkelte skole. Kortlægning og analyse skal som minimum forholde sig til:
- Hvordan og i hvilket omfang der politisk træffes beslutninger på baggrund af de nationale test.
 - Hvad man politisk og forvaltningsmæssigt supplerer med for at have sikker viden om skolernes faglige status.
 - En kortlægning af, hvorvidt man på kommunalt niveau anvender muligheden for at følge faglig progression gennem hele skoleforløbet.
15. *Hvordan anvendes de nationale test på nationalt styringsniveau?*
Der skal gennemføres en kortlægning og analyse på nationalt niveau af anvendelse af viden fra de nationale test i relation til politisk beslutningsgrundlag, tilsyn og andre indsatser samt forskning på skoleområdet. Kortlægning og analyse skal som minimum forholde sig til:
- En kortlægning af, hvordan og i hvilket omfang viden fra de nationale test anvendes i forhold til det politiske niveau.
 - En kortlægning af, hvordan og i hvilket omfang viden fra de nationale test anvendes i viden og forskning, der anvendes af det politiske niveau.
 - Om der kan indhentes mål for faglighed af anden vej om elevernes skoleforløb over tid frem mod folkeskolens afgangsprøve.
 - En kortlægning af, hvorvidt man på nationalt niveau anvender muligheden for at følge faglig progression gennem hele skoleforløbet.

Delopgavens metode: Evalueringsopgaven omkring de nationale test skal give viden på tre niveauer: 1) målopfyldelse, 2) implementering og 3) som teoribaseret evaluering.

Målopfyldelsesevaluering

Der skal foretages en evaluering af graden af hvorvidt de fastsatte mål for de nationale test er opnået, og i hvilken grad de nationale test har bidraget til opfyldelsen af de formål der blev givet for de nationale test.

Implementeringsevaluering

Som en del af evalueringen af de nationale test skal der foretages en evaluering af implementeringen af testene. Evalueringen af implementeringen skal undersøge hvilke faktorer, der har fremmet eller besværliggjort implementeringen samt beskrive de involverede aktørers (elever, lærere, ledere, forvaltnings- og politiske aktører) erfaringer med de nationale test, samt de fordele og ulemper, de ser ved brugen af de nationale test.

Teoribaseret evaluering og metoder

Den teoribaserede evaluering har til formål at undersøge, om de forventede resultater ved brugen af de nationale test i forhold til de givne formål er opnået. Den teoribaserede evaluering skal derfor bero på et evalueringsdesign, som muliggør undersøgelse af sammenhængen mellem indsatserne med de nationale test og resultaterne. Evalueringsdesignet skal gøre det muligt at identificere de forhold og de mekanismer, der synliggør hvordan de nationale test virker for de respektive aktører.

Der skal indgå kvantitative og kvalitative dataindsamlinger i evalueringen i forhold til relevante forhold, der gør sig gældende for aktører og aktiviteter omkring brugen af de nationale test. Det kan eksempelvis være spørgeskemaundersøgelse, anvendelse af registerdata, forløbs- og testdata, brug af data fra følgeforskningen til folkeskolereformen, kvalitative casestudier og deskresearch. Det er et krav, at der gennemføres tiltag til at validere resultater af undersøgelsen. Det kan være ved kvalificeringsworkshop m.v. Evaluator kan komme med yderligere forslag til undersøgelsesdesign.

Evaluator skal sandsynliggøre, at undersøgelsesdesignet giver solide analyser. Evaluator skal endvidere tilbyde et design, der sikrer repræsentativitet i forhold til demografisk og geografisk spredning og variation i kommune- og skolestørrelser m.v.

Formidling og endeligt produkt

Evaluator skal foretage en samlet afrapportering af delopgave et og to. Rapporten skal afveje evalueringens resultater for de respektive aktører op mod formålene for de nationale test og give mulighed for, at der kan anvises retning for den fremadrettede udvikling og brug i folkeskolen.

Tidsmæssige krav

Det forudsættes, at der er en tæt dialog mellem Styrelsen for Undervisning og Kvalitet og evaluator med månedlige statusmøder.

Evaluering og afrapportering udarbejdes efter følgende tidsplan:

- Opstartsmøde primo marts 2019
- Dataindsamling i forbindelse med dem obligatoriske testafvikling fra 1. marts til 30. april.
- Statusnotat ultimo august som opfølgning på dataindsamlingen og de indledende analyser.
- Udkast til rapport afleveres til STUK ultimo oktober 2019
- Endelig rapport afleveres til STUK ultimo november 2019
- STUK offentliggør rapport med udgangen af januar 2020

Budget

Opgaven skal løses inden for en ramme på 3.000.000 kr., der faktureres med aflevering af den endelige rapport.