

Ekspertgruppen for Folkeskolens Nationale Færdighedstest

Anbefalinger, september 2023

Indhold

1	Indledning	3
1.1	Baggrund	3
1.2	Ekspertgruppens kommissorium og arbejde	3
1.3	Ekspertgruppens sammensætning	4
2	Anbefalinger	5
3	Ekspertgruppen anbefaler	7
3.1.	Fagligt indhold	7
3.2.	Testresultaternes anvendelse	8
3.3.	Psykometriske egenskaber	11
3.4.	Testafholdelsen	16
4	Et flertal i ekspertgruppen anbefaler	20
4.1.	Fagligt indhold	20
4.2.	Testresultaternes anvendelse	25
4.3.	Psykometriske egenskaber	34
4.4.	Testafholdelsen	45
4.5.	Øvrige temaer	48
5	Et mindretal i ekspertgruppen anbefaler	51
5.1.	Fagligt indhold	51
5.2.	Testresultaternes anvendelse	52
5.3.	Øvrige temaer	53
6	Bilag	55

1 Indledning

1.1 Baggrund

Folkeskoleforligskredsen indgik den 29. oktober 2021 *aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen*. Det følger af aftalen, at Folkeskolens Nationale Færdighedstest skal erstatte de tidligere nationale test og de nuværende Folkeskolens Nationale Overgangstest. Folkeskolens Nationale Færdighedstest skal være klar til skoleåret 2026/27.

For Folkeskolens Nationale Færdighedstest gælder blandt andet, at:

- Testene skal være lineære, så alle elever får de samme opgaver.
- Testene skal gennemføres i starten af skoleåret, så de kan bidrage til tilrettelæggelsen af undervisningen resten af skoleåret.
- Testene skal fokusere på centrale færdigheder for elevers skolegang i læsning og matematik.
- Testene skal anvendes hyppigt i det samme fag, så de kan styrke den faglige udvikling.
- Testene skal være digitalt selvrettende.

Ud over de nye Folkeskolens Nationale Færdighedstest skal der udvikles en række nye frivillige test, som også vil blive stillet til rådighed for skolerne i skoleåret 2026/27.

Med *aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen* blev det besluttet, at faglige eksperter og *Sammen om Skolen* inddrages i udvikling og implementering af Folkeskolens Nationale Færdighedstest.¹ Børne- og undervisningsministeren nedsatte i marts 2023 en ekspertgruppe, som skulle give faglige anbefalinger for rammerne for Folkeskolens Nationale Færdighedstest. Ud over ekspertgruppen inddrages *Sammen om Skolen* løbende via en følgegruppe, hvor repræsentanter fra *Sammen om Skolen* kan give deres input til arbejdet med testene.

1.2 Ekspertgruppens kommissorium og arbejde

Ekspertgruppens arbejde har taget udgangspunkt i et kommissorium udarbejdet af Børne- og Undervisningsministeriet.²

Kommissoriet giver ekspertgruppen følgende formål:

Ekspertgruppen skal bidrage til at sikre et kvalificeret og solidt fagligt grundlag for den kommende udvikling og brug af Folkeskolens Nationale Færdighedstest og øvrige test. Det er afgørende, at anbefalingerne tager højde for de forskellige faglige perspektiver i de relevante fag og discipliner, herunder sikrer, at de overordnede rammer for test i de forskellige fag så vidt muligt er de samme. Ekspertgruppens anbefalinger vil indgå i ministeriets videre arbejde med at udvikle testene.

¹ Børne- og Undervisningsministeriet. (2021). *Bred aftale om fremtidigt evaluerings- og bedømmelsessystem*: <https://www.uvm.dk/aktuelt/nyheder/uvm/2021/okt/211029-bred-aftale-om-fremtidigt-evaluerings-og-bedømmelsessystem>

² Børne- og Undervisningsministeriet. (2023). *Kommissorium for ekspertgruppe for Folkeskolens Nationale Færdighedstest og øvrige test*: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf23/marts/230307-kommissorium-for-ekspertgruppe-for-folkeskolens-nationale-faerdighedstest-og-oevrige-test.pdf>

På fem møder fra marts til juni 2023 har gruppens medlemmer udarbejdet og diskuteret anbefalinger for rammerne for Folkeskolens Nationale Færdighedstest.

1.3 Ekspertgruppens sammensætning

Det fremgår af *kommissoriet for ekspertgruppe for Folkeskolens Nationale Færdighedstest og øvrige test*, at ekspertgruppen består af 18 personer. Foruden formanden er ekspertgruppen sammensat af lærere, en skoleleder, en forvaltningsrepræsentant, samt forskere, der har viden om test, evaluering, psykometri og fagdidaktik, og som kommer fra forskellige forsknings- og vidensmiljøer, herunder professionshøjskolerne. Derudover var der udpeget en elevrepræsentant, som trådte ud af gruppen, da vedkommende ikke havde mulighed for at deltage i arbejdet. Ekspertgruppen har derfor bestået af 17 medlemmer.

Medlemmerne er:

Professor Peter Dahler-Larsen (formand), Københavns Universitet

Docent Charlotte Krog Skott, Professionshøjskolen Absalon

Lektor Kaj Østergaard, VIA Læreruddannelsen

Lektor Mads Poulsen, Københavns Universitet

Lektor Tine Nielsen, UCL Erhvervsakademi og Professionshøjskole

Lærer Adina Lynn Raskin Teplin, Gribskov Kommune

Lærer Emil Hjerl, Gribskov Kommune

Lærer Hanne Sax Holm, Vejen Kommune

Lærer Maiken Adelsten, Bornholms Regionskommune

Lærer Marie Bonde Clausen, Hvidovre Kommune

Lærer Søren Nørgaard Meinert, Holbæk Kommune

Skolechef Mette Rønø, Lyngby-Taarbæk Kommune

Skoleleder Thomas Dandanell Nielsen, Faxe Kommune

Professor emeritus Hanne Foss Hansen

Professor Jeppe Bundsgaard, Aarhus Universitet

Professor Rolf Vegar Olsen, Universitetet i Oslo

Professor Simon Calmar Andersen, Aarhus Universitet

2 anbefalinger

Det følger af kommissoriet, at ekspertgruppen som minimum skal udarbejde anbefalinger i en samlet rapport inden for følgende fire temaer:

- Tema 1: Rammer for testens faglige indhold
- Tema 2: Rammer for testresultaternes anvendelse
- Tema 3: Rammer for testens psykometriske egenskaber
- Tema 4: Rammer for testens afholdelse

Derudover har ekspertgruppen også haft mulighed for at udarbejde anbefalinger til øvrige temaer.

Ved læsning af anbefalingerne bør man være opmærksom på, at temaerne har indbyrdes sammenhænge, og at der i mange tilfælde vil være gensidige afhængigheder. Beslutninger relateret til et tema vil således påvirke beslutninger relateret til øvrige temaer. Temaerne bør derfor ikke behandles isoleret, men bør ses i sammenhæng.

Ifølge kommissoriet skal anbefalingerne så vidt muligt afspejle enighed, men der er plads til meningsforskelle. Anbefalingerne skal ligeledes tage højde for de givne formål og kriterier ved Folkeskolens Nationale Færdighedstest, samt tage højde for temaernes indbyrdes afhængighed.

Det har ekspertgruppen fortolket således, at indstillingen fra ekspertgruppen skal give plads til både enighed og meningsforskelle. Ligeledes har det været ekspertgruppens ønske at afspejle de argumenter for og imod en given anbefaling, som har været særligt afgørende for de enkelte medlemmers stillingtagen, for på den måde at give eksempler på de afvejninger, der ofte må præge en efterfølgende beslutning.

Følgende spilleregler har været gældende i ekspertgruppens arbejde: Hvert enkelt medlem har haft mulighed for at fremsætte forslag til anbefalinger. Alle modtagne forslag er blevet drøftet. Ethvert forslag til anbefalinger er taget med i indstillingen, medmindre det med forslagsstillerens accept er trukket tilbage eller skrevet sammen med et andet forslag.

Det har medført, at anbefalinger om samme emne med varierende sprogbrug og forskellige grader af overlap eller modstrid alle er bragt til afstemning. Når man læser disse anbefalinger efter hinanden, kan det være en fordel at have in mente, at forskellige flertal i ekspertgruppen kan have stillet sig bag de forskellige varianter. Som konsekvens af den valgte arbejdsform i gruppen er det fundet fair og redeligt at videregive alle de forslag for læseren, om hvilke ekspertgruppens medlemmer har givet deres meninger til kende.

For hver enkelt anbefaling har det enkelte medlem haft mulighed for at støtte anbefalingen eller ikke at støtte anbefalingen. Det har ligeledes været muligt at forholde sig neutralt til den enkelte anbefaling. Begrundelser herfor kan eksempelvis være, at man ikke har følt sig fagligt rustet på det specifikke område, som anbefalingen omhandler, at man kan lade anbefalingen passere med de argumenter for og imod, der er opstillet, eller at man har fundet, at stillingtagen til anbefalingen i sidste instans bør være en politisk afgørelse.

I lyset af medlemmernes stillingtagen og af ekspertgruppens drøftelser er anbefalingerne blevet justeret løbende, og der er givet mulighed for ny stillingtagen i flere runder. Deadline for den endelige meningstilkendegivelse var den 27. juni 2023 klokken 09.00.

En anbefaling er afslutningsvis placeret under overskriften "**Ekspertgruppen anbefaler**", hvis ingen medlemmer har ønsket ikke at tilslutte sig en anbefaling.

En anbefaling er afslutningsvis placeret under overskriften "**Et flertal i ekspertgruppen anbefaler**", hvis der blandt de medlemmer, der har ønsket at stemme om en anbefaling, er flere for end imod anbefalingen.

En anbefaling er afslutningsvis placeret under overskriften "**Et mindretal i ekspertgruppen anbefaler**", hvis der blandt de medlemmer, der har ønsket at stemme om en anbefaling, ikke er et flertal for anbefalingen.

Kategoriseringen af en anbefaling i forhold til flertal/mindretal kan således være baseret på forskellige antal afgivne stemmer. For så vidt det gælder *anbefaling nr. 22* og *anbefaling nr. 23* har færre end halvdelen af ekspertgruppens medlemmer ønsket at afgive deres stemme. Opgørelsen af flertal/mindretal bør ses i det lys.

Afstemningen gælder anbefalingens overskrift, inklusive den forklarende tekst, som er skrevet eller godkendt af anbefalingens forfatter(e).

Argumenter for at følge eller ikke følge en given anbefaling kan være fremsat af ethvert medlem af ekspertgruppen. De er gengivet i rapporten for at give læseren et indblik i aspekter af gruppens drøftelser. Argumenter for og imod en anbefaling udgør eksempler på argumenter, men skal ikke forveksles med en fuldstændig analyse. At argumenterne er gengivet i indstillingen betyder ikke, at ekspertgruppen som helhed har godkendt hvert enkelt af de gengivne argumenter. Det enkelte medlem har ikke haft mulighed for at fjerne en anbefaling eller et argument for eller imod en anbefaling, som et andet medlem har fremsat.

Ekspertgruppens medlemmer har haft mulighed for ved navns nævnelser i fodnoter at give udtryk for, at de støtter henholdsvis ikke støtter den enkelte anbefaling.

Rækkefølgen og nummereringen af de opstillede anbefalinger er ikke udtryk for en prioritering af deres relative vigtighed.

I tilgift til de opdelte anbefalinger har ekspertgruppens medlemmer haft mulighed for under eget forfatterskab at skrive såkaldte positionspapirer. Et positionspapir giver fx mulighed for en længere sammenhængende fremstilling af begrundelser for en given stillingtagen til de forskellige anbefalinger. Der er fremkommet to positionspapirer (se bilag).

3 Ekspertgruppen anbefaler

3.1. Fagligt indhold

ANBEFALING 1

Det anbefales, at der skal være en tydelig sammenhæng mellem Folkeskolens Nationale Færdighedstest og de øvrige nationale styringsredskaber i forhold til fagligt indhold og didaktisk tilgang.

Folkeskolens Nationale Færdighedstest er tænkt som et nationalt redskab på linje med de øvrige nationale styringsredskaber som faghæfter (Fælles Mål og læseplaner) og folkeskolens prøver. En tydelig sammenhæng mellem disse styringsredskaber kan derfor bidrage med en vision og retning for folkeskolens undervisning, som støtter de samme faglige og didaktiske prioriteringer. Indholdet af Folkeskolens Nationale Færdighedstest skal derfor passe til fagets fagformål, kompetencemål, underliggende færdigheds- og vidensområder, samt de vejledende færdigheds- og vidensmål (som beskrevet i faghæfter) og det skal passe til opgavetyper og fordelingen af dem i folkeskolens prøver.

Argumenter for at følge

- I. Hvis anbefalingen ikke følges og Folkeskolens Nationale Færdighedstest tester et begrænset antal færdigheder, så er der stor risiko for: at undervisningens indhold centrerer sig om disse færdigheder, og at undervisningens didaktiske tilgang bliver mere traditionel (dvs. høj grad af lærerstyring, gennemgang af fagligt stof og træning af færdigheder, modsat undersøgende elevaktiviteter, faglige dialoger og refleksioner, som fremhæves i faghæfter). Denne risiko for *teaching to the test* underbygges af forskning³, der fremhæver det som en effekt af test, prøver og eksamener. Risikoen bliver ikke mindre, når omfanget af de obligatoriske test tages i betragtning.
- II. Hvis anbefalingen ikke følges og Folkeskolens Nationale Færdighedstest tester et antal færdigheder, der både er indeholdt i og ikke-indeholdt i faghæfterne, så er der risiko for, at Folkeskolens Nationale Færdighedstest kommer til at udgøre implicite standarder for fagene, som det fx er sket i Norge (Seland & Hovdhaugen, 2017⁴).
- III. Argumenterne I og II kan ses som et samlet argument for at undgå, at Folkeskolens Nationale Færdighedstest kommer til at virke mod hensigterne i andre styringsredskaber.
- IV. Folkeskolens Nationale Færdighedstest vil blive brugt til udsagn om klassers, folkeskolers og folkeskolesystemets status og udvikling. Udsagn på nationalt og kommunalt plan og på skoleplan om, hvordan det går med elevernes faglige niveau, har stor opmærksomhed. Det er derfor afgørende, at Folkeskolens Nationale Færdighedstest måler så bredt, at resultaterne reelt

³ Barnes, M., Clarke, D.J., & Stephens, M. (2000). *Assessment: The engine of systematic curricular reform?* Journal of Curriculum Studies, 32 (5), s. 623-650.

⁴ Selan, I., & Hovdhaugen, E. (2017). *National Tests in Norway: An Undeclared Standard in Education? Practical and Political Implications of Norm-Referenced Standards*. I: Sigrid Blömeke og Jan-Eric Gustafsson (eds.) Standard setting in education. The Nordic countries in an international perspective, s. 161-180. Springer.

er udtryk for det faglige niveau, og ikke blot for et afgrænset udsnit af fagene eller de faglige områder, der testes.

- V. Det er vigtigt, at offentligheden og ledelser har viden om faget/fagområdet som sådan og ikke et løsrevet udsnit.

Argumenter for ikke at følge

- I. Dele af kompetenceområderne er vanskelige at teste i en digital selvrettende test. *Anbefaling nr. 75* kan imødekomme dette modargument.

3.2. Testresultaternes anvendelse

ANBEFALING 2

Det anbefales, at lærerne klædes testfagligt tilstrækkeligt på til at kunne varetage testsituationen på en ensartet måde, samt klædes på til at give fyldestående tilbagemelding til den enkelte elev og samtale med forældre omkring testresultater. Det anbefales desuden, at skoleledere og skoleansvarlige i kommuner klædes testfagligt på til at kunne understøtte lærernes arbejde.⁵

Lærere er fagprofessionelle, som pålægges opgaven med:

1. at varetage selve testningen af eleverne med alle de situationer, der kan opstå,
2. at give individuel tilbagemelding og arbejde pædagogisk på grundlag af testresultater, samt
3. at samtale med forældre om deres børns testresultater.

På baggrund af Bundsgaard & Pucks (2016)⁶ undersøgelse af blandt andet læreres forståelse af resultaterne fra de tidligere nationale test i folkeskolen, bør det sikres, at lærerne er klædt testfagligt tilstrækkeligt på til at udføre disse opgaver. Da det desuden forventes, at formatet på tilbagemeldingerne af Folkeskolens Nationale Færdighedstest kommer til at fremstå anderledes end tidligere, givet nyt test-design m.v., er der et behov for at løse dette fremadrettet.

Løsningen kan indeholde flere elementer, herunder vejledninger om Folkeskolens Nationale Færdighedstest til lærere og forældre, vejledninger om Folkeskolens Nationale Færdighedstest generelt, og andre eventuelle støttematerialer, som bør stilles til rådighed fra Børne- og Undervisningsministeriets side (jf. *anbefaling nr. 38*). En del af løsningen kan også være, at lærerstuderende - med undervisningsfag som også er fag, hvor der testes i Folkeskolens Nationale Færdighedstest - undervises i testvaretagelse, testtilbagemelding og testforståelse, som en del af det relevante undervisningsfag eller fx i sidste semester af læreruddannelsen, hvor dette også kan afprøves i den afsluttende praktik.

Argumenter for at følge

- I. Øget forståelse omkring testresultater og hvad resultaterne betyder og øget tilbagemeldingskompetence. Samlet vil det også bidrage til øget kvalitet i lærernes arbejde med den enkelte elev på grundlag af testresultater, samt øget kvalitet i samtaler med forældre om deres børns testresultater.

⁵ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 39* og *48*.

⁶ Bundsgaard J. & Puck M. R. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet. Center for Anvendt Skoleforskning ved University College Lillebælt: <https://dpu.au.dk/fileadmin/edu/Forskning/Ebog - Nationale test - FINAL - september 2016.pdf>

ANBEFALING 3

Det anbefales, at Børne- og Undervisningsministeriet udarbejder en vejledning til lærere, der støtter dem i at anvende testresultaterne til at tilrettelægge den videre undervisning med fokus på elevernes faglige udvikling.

Lærerne skal støttes i at bruge testens resultater konstruktivt og meningsfuldt både på elev- og klasse-niveau. I forbindelse med hver test bør der derfor udarbejdes en vejledning til lærerne, så de støttes i at anvende testresultaterne til at tilrettelægge undervisningen "med fokus på elevernes varierede udgangspunkter og sikrer deres faglige udvikling" (jf. *den politiske aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen*). Vejledningen skal blandt andet indeholde:

- Generelle intentioner med testen eventuelt set i sammenhæng med test fra tidligere år og test fra andre klassetrin.
- En overordnet oversigt over testens opgavetyper og indhold, herunder fx fordeling mellem stofområder og eventuelt udvalgte fokusområder.
- Oversigt over, hvilke konkrete elementer fra læseplanen hver enkelt testitem tester, og hvordan de enkelte testitems kan ses i sammenhæng.
- Estimerede sværhedsgrader af de enkelte testitems.
- Typiske fejl og sammenhæng med typiske misopfattelser.
- Muligheder for tiltag ved udvalgte elevsvar både på elev- og klasseniveau.

Argumenter for at følge

- I. En vejledning vil støtte lærerne til at bruge testresultaterne på en valid måde.
- II. Vejledningen kan medvirke til at sikre transparens og sammenhæng mellem testudviklernes intentioner og den måde testene bruges på i skolen.
- III. Vejledningen kan hjælpe lærerne til at føle sig sikre i brugen af testresultaterne og dermed øge den konkrete brug i skolen.

ANBEFALING 4

Det anbefales, at Børne- og Undervisningsministeriets nuværende kriterier og praksis for ansøgning om og udlevering af aggregerede data fra test både på elev, klasse, skole og kommuneniveau, (samt med og uden person-identifikation) videreføres til Folkeskolens Nationale Færdighedstest. Det anbefales endvidere, at praksissen udvides således, at der også udleveres sumscores, samt elementære oplysninger vedrørende kvaliteten og måleusikkerheden af alle scores i Folkeskolens Nationale Færdighedstest.⁷

For data fra de tidligere og nuværende nationale test i folkeskolen findes retningslinjer for ansøgning om og udlevering af aggregerede data⁸. Disse retningslinjer angiver, at det er muligt at ansøge om at få udleveret aggregerede data på elev, klasse, skole og kommuneniveau i form af Rasch-scores, percentil-scores, og kriterie-scores, samt at disse kan søges udleveret med og uden person-identifikation, hvor førstnævnte muliggør kobling til relevante registerdata og analyser via Danmarks Statistisk forskerordning eller selvstændige analyser. Udleveringen af de tre typer scores bør bibeholdes.

⁷ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 5*.

⁸ Børne- og Undervisningsministeriet. (2016). *Retningslinjer for brug af data fra de nationale test til statistiske og videnskabelige undersøgelser*: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf16/mar/160315-retningslinjer-for-brug-af-data-fra-de-nationale-test-til-statistik-og-videnskab.pdf>

Argumenter for at følge

- I. Vil muliggøre, at der gennemføres nødvendig og uafhængig forskning i gruppeforskelle og tidlig udvikling i resultater fra Folkeskolens Nationale Færdighedstest, samt deres sammenhæng med andre centrale variable med brug af en flerhed af metoder og med udgangspunkt i forskellige fagligt funderede hypoteser og forskningsspørgsmål.

ANBEFALING 5

Det anbefales, at det i retningslinjer for ansøgning om og udlevering af data fra Folkeskolens Nationale Færdighedstest (jf. anbefaling nr. 4) også sikres, at der udleveres item-data på elevniveau, inklusiv de af Styrelsen for It og Læring beregnede scores og elementære nødvendige informationer. Det anbefales endvidere, at udleveringen kan ske med eller uden person-identifikation, for så vidt et legitimt formål kan begrundes.⁹

Det er vigtigt, at uafhængige forskere kan replikere Børne- og Undervisningsministeriets psykometriske analyser af data fra Folkeskolens Nationale Færdighedstest, og dermed også opdage eventuelle svagheder i de gennemførte analyser, således som det var tilfældet med de tidligere nationale test i folkeskolen. Bundgaard og Kreiners analyser af data fra de adaptive nationale test udgjorde fx en del af startpunktet til, hvad der endte med at være en politisk beslutning om at gå bort fra et adaptivt testdesign og i stedet lave et nyt testsystem med et lineært testdesign i de nye Folkeskolens Nationale Færdighedstest, som ekspertgruppen fremsætter de nærværende anbefalinger om¹⁰.

Det er også væsentligt, at der kan gennemføres analyser med brug af andre psykometriske metoder og anden statistisk software, end det som anvendes i Børne- og Undervisningsministeriet, således at der kan produceres yderligere viden om, hvordan Folkeskolens Nationale Færdighedstest fungerer målingsmæssigt.

Til brug for *differentiel item funktion*-analyser (DIF-analyser) bør data på item-niveau kunne udleveres både med og uden person-identifikation, medfølgende de relevante baggrundsvariable, som Børne- og Undervisningsministeriet er i besiddelse af. Det er helt centralt med baggrundsvariable, således at psykometriske analyser kan gennemføres uden om Danmarks Statistik, da adgangen til specialiserede softwarepakker til psykometriske analyser er yderst begrænset på Danmarks Statistiks forskerordning. Udlevering bør, ud over data på item-niveau og baggrundsvariable suppleres med de forskellige typer af scores, der udleveres bredere (jf. anbefaling nr. 4). Endelig bør en dataudlevering til psykometriske formål suppleres med opgaverne og deres relation til datasættet, således at det er muligt at forstå og tolke resultaterne.

⁹ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 4*.

¹⁰ Bundgaard, J. & Kreiner, S. (2019). *Undersøgelse af De Nationale Tests måleegenskaber*. DPU, Aarhus Universitet.

¹⁰ Styrelsen for It og Læring (STIL). (2020). *Evaluering af de statistiske aspekter ved de nationale test*. STIL. Børne- og Undervisningsministeriet.

¹⁰ Det Nationale Forsknings- og Analysecenter for Velfærd, VIVE. (2020). *Review af evalueringen af de statistiske aspekter ved de nationale test – Delrapport 1: Evaluering af de nationale test*. VIVE.

¹⁰ Nielsen, T. (2019). *Review af evalueringen af de statistiske aspekter ved de nationale test – Delrapport 1: Evaluering af de nationale test*: [Microsoft Word - REVIEW AF DOKUMENTATIONEN FOR DE NATIONALE TEST TINE NIELSEN3.docx](#)

Argumenter for at følge

- I. Det vil være et stærkt politisk signal at give mulighed for, at Børne- og Undervisningsministeriets psykometriske analyser kan replikeres, efterprøves og på anden vis valideres ved at give mulighed for, at der i forbindelse med Folkeskolens Nationale Færdighedstest kan laves uafhængig psykometrisk forskning på opgavernes psykometriske egenskaber, linkingen over klassetrin og tid m.v. Baggrund er, at dette tidligere har vist sig nyttigt og har bidraget til den nuværende udvikling.
- II. Det vil samtidig give mulighed for, at der gennemføres psykometrisk forskning ved hjælp af anden psykometrisk software, end hvad der anvendes af Styrelsen for It og Læring. Denne forskning kan potentielt også bringe Styrelsen for It og Læring videre.

3.3. Psykometriske egenskaber

ANBEFALING 6

Det anbefales, at det tilstræbes at udvikle tidssvarende og interaktive testopgaver til Folkeskolens Nationale Færdighedstest.

Det betyder noget for elevernes testoplevelse, om opgaverne er kedelige og ensartede, eller om de er smukt designede og lægger op til interaktion. Det betyder også noget, om testopgaverne er af en form, hvor eleverne kan få oplevelsen af, at de ikke kan finde ud af det, eller om de kan løse opgaven på deres niveau, på en måde så de ikke oplever nederlag. Derfor skal det tilstræbes at udvikle grafisk tidssvarende, interaktive og ikke-simple opgaver. Interaktive opgaver kan fx indebære, at eleven flytter rundt på objekter på skærmen, visuelt programmerer en robot, laver udregninger i et regneark, designer 3D-figurer og meget mere.

Argumenter for at følge

- I. De gamle nationale test gav mange elever en dårlig oplevelse. Ifølge lærerne i en undersøgelse fra 2016 (Bundsgaard & Puck 2016¹¹) var det selv blandt de dygtigste elever under hver femte, der i høj eller meget høj grad havde en god oplevelse med testen.
- II. Elever, der begynder at kede sig under testafviklingen, vil være mindre fokuserede på at svare det bedste de kan, hvilket vil give mindre korrekte resultater.
- III. Med interaktive opgaver kan en langt bredere del af de faglige færdigheder udprøves.
- IV. Muligheden for at skabe multimodale og animerede opgaver gør, at man kan reducere mængden af forklarende tekst og derved fokusere på den faglige færdighed.
- V. Der er en lang tradition i internationale undersøgelser (PISA, ICILS, PIRLS, TIMSS) for at udvikle avancerede opgaver, som de her omtalte. Der er således ikke tale om eksperimentel udvikling, men om at bygge videre på velafprøvede formater.
- VI. Opgaverne skal som altid afprøves af elever, inden de tages i brug. Derved kan det sikres, at de interaktive funktioner forstås og fungerer som ønsket.

¹¹ Bundsgaard J. & Puck M. R. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet. Center for Anvendt Skoleforskning ved University College Lillebælt: <https://dpu.au.dk/fileadmin/edu/Forskning/Ebog - Nationale test - FINAL - september 2016.pdf>

ANBEFALING 7

Det anbefales, at Folkeskolens Nationale Færdighedstest designes, så de så vidt muligt både indeholder opgavetyper, der er dikotome og polytome eller testlets.¹²

Klassiske evne- og færdighedstest indeholder typisk opgaver, der kan besvares rigtigt eller forkert. Det kan være i form af såkaldte dikotome opgaver, hvor der kun er én rigtig og én forkert svarmulighed, eller det kan være i form af multiple choice-opgaver, som har et rigtigt og flere forkerte svar. Sådanne opgaver gør det let at have et selvrettende testsystem, da det muliggør, at man i forvejen har kodet hvilken svarmulighed, der er den rigtige og som dermed bidrager til elevens score (forkerte svar scores 0 og korrekte svar scores 1). Derudover er det muligt at have ægte polytome eller testlet opgaver af forskellig slags, hvor der kan gives flere point per opgave, eller hvor nogle svar giver delvise point. Alle disse typer af opgaver kan der udvikles objektive scoringsnøgler til. Opgavescoringsnøglerne kan implementeres direkte ind i testsystemet, som således bliver selvrettende.

Der er klare fordele ved at inkludere opgavetyper, der er polytome eller testlet opgaver, i Folkeskolens Nationale Færdighedstest, idet disse opgavetyper, hvis de er godt konstrueret, giver mere information, end de indeholdte dikotome opgaver ville gøre, hvis de var enkeltstående opgaver. Polytome og testlet-opgaver vil, ved at bidrage med mere information, alt andet lige bidrage til mindre måleusikkerhed/større målepræcision med et mindre antal opgaver, end hvis der udelukkende anvendes rent dikotome og adskilte opgaver. Dermed kan testtiden også bringes ned, samtidigt med, at måleusikkerheden er lav (jf. *anbefaling nr. 14* og *anbefaling nr. 45*).

Argumenter for at følge

1. Det er ønskværdigt, at Folkeskolens Nationale Færdighedstest indeholder opgavetyper, som både kan være selvrettende og bidrager til at nedbringe både testtiden og måleusikkerheden, således at eleverne ikke testes unødigt længe og de opnår testresultater med den præcision, som er krævet for at gennemføre individuel personvurdering ud fra internationale, herunder europæiske, standarder for pædagogiske test.

ANBEFALING 8

Det anbefales, at kvaliteten af multiple choice-opgaver, og særligt distraktorerne sikres.

Folkeskolens Nationale Færdighedstest skal i højere grad lægge vægt på, at opgaverne i sig selv skal give læreren nyttig information til brug for opfølgning med eleverne. Ud fra en antagelse om, at der kommer til at være et stort antal multiple choice-opgaver, beskrives her nogle nærmere principper for afprøvning, som vil være relevante at følge for at teste, om også distraktorerne giver meningsfuld didaktisk information:

1. Første afprøvning: Er afprøvning med henblik på at identificere meningsfulde distraktorer. Opgaven gives som et åbent spørgsmål. Elevernes svar kategoriseres, og højfrekvente svar benyttes som distraktorer i konstruktionen af multiple choice-version.
2. Anden afprøvning: Fokusgruppe med elever efter såkaldte "tænke-højt"-procedurer. Fx kan to elever arbejde sammen, mens de taler højt under løsningen af opgaven. Hensigten med denne afprøvning er at forsikre sig om, at distraktorerne repræsenterer de forestillinger, som man på forhånd har tænkt, at de skal repræsentere.

¹² Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 8, 9, 10, 45, 46, 50, 51* og *positionspapir fra Rolf Vegar Olsen med flere*.

3. Tredje afprøvning: Er den store psykometriske afprøvning af alle opgaver (jf. *anbefaling nr. 10*). I denne skal der typisk være to til tre gange så mange opgaver, som den endelige test skal have.

Argumenter for at følge

- I. Dette skal gøres for, at opgaverne i højere grad kan blive et pædagogisk værktøj.
- II. En sådan procedure anvendes i andre testsystemer, og har vist sig at lede til flere såkaldte effektive distraktorer.
- III. En sådan procedure vil også bidrage til god begrebsvaliditet (jf. *anbefaling nr. 46*).
- IV. En procedure, der desuden omfatter analyser af Differentiel Distraktor Funktion, vil desuden kunne give information i forhold til, om der er systematiske misopfattelser hos grupper af elever i forhold til bestemte forkerte svar, og dette kan således undgås - på samme måde som *differentiel item funktion* bør undgås (jf. *anbefaling nr. 10*).

Argumenter for ikke at følge

- I. Udvidede rutiner for afprøvning af opgaver giver ekstra udgifter.

ANBEFALING 9

Det anbefales, at der gennemføres grundige indholdsmæssige og formatmæssige afprøvningsopgaver til Folkeskolens Nationale Færdighedstest, og at afprøvningsopgaverne kvalitetssikres.¹³

Udvikling af testopgaver er en omfattende og krævende proces, som kræver både fagfaglig og testfaglig ekspertise. Opgaverne bør vurderes på mindst tre forskellige aspekter og af personer uden for udviklingsgruppen, inden der indsamles data til psykometrisk afprøvning og vurdering.

De tre forskellige aspekter er:

1. Indholdsmæssig vurdering af uafhængige fagekspertter, dvs. indholdsvaliditetsvurdering - dækker opgaverne det færdighedsområde, der ønskes målt godt nok, og kun det færdighedsområde.
2. Forståelsesmæssig vurdering, dvs. at elevmålgruppen forstår opgaverne.
3. Format- og teknisk opsætningsmæssig vurdering, dvs. at eleverne let og intuitivt kan besvare diverse opgaveformater, uden at skulle bruge tid på at sætte sig ind i selve måden, der skal svares på, og således at der ikke måles fx teknisk literacy i tillæg. Derudover bør en uafhængig kvalitetssikring følge processen og vurdere resultaterne af denne afprøvning af opgaverne (jf. *anbefaling nr. 12*).

Argumenter for at følge

- I. En grundig afprøvning af opgaverne indholds- og formatmæssigt vil bidrage til begrebsvaliditeten, oplevelsen af relevans og målesikkerheden af de enkelte test.
- II. Uafhængig kvalitetssikring vil bidrage til at højne kvaliteten af de endelige Folkeskolens Nationale Færdighedstest.

¹³ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 10, 15, 17, 46* og *positionsrapport fra Rolf Vegar Olsen med flere*.

ANBEFALING 10

Det anbefales, at der gennemføres grundige psykometriske afprøvninger af opgaverne og færdighedsskalaerne i de nye lineære Folkeskolens Nationale Færdighedstest. Det anbefales også, at de psykometriske kvalitetskrav til Folkeskolens Nationale Færdighedstest er højere end ved de tidligere adaptive nationale test, og at de psykometriske afprøvninger kvalitetssikres.¹⁴

Det er essentielt, at alle opgavesæt i Folkeskolens Nationale Færdighedstest undersøges og kvalitetssikres psykometrisk med højere kvalitetskrav end de, som blev anvendt på de tidligere adaptive nationale test. De psykometriske kvalitetskrav bør omfatte nedenstående almene krav:

- Hver skala, som giver en score i sig selv, skal kun måle en begrebskonstruktion/et færdighedsområde (psykometrisk krav om unidimensionalitet).
- Opgavesvarene skal være indbyrdes positivt korrelerede, og svar på de enkelte opgaver skal være korrelerede til scores (psykometrisk krav om konsistens og monotonicitet).
- Opgaverne skal fungere på samme måde for relevante undergrupper af elever, fx henover køn og modersmålsbaggrund for, at der er sammenlignelighed af scores henover undergrupperinger af elever (psykometrisk krav om invarians og ingen *differentiel item funktion*).
- Opgavesvar skal kun afhænge af niveauet af den færdighed, der måles, og ikke af svarene på anden/andre opgave(r) (psykometrisk krav om lokal uafhængighed).
- Det skal desuden sikres, at der er god tilpasning mellem opgavernes sværhedsgrader og elevernes færdighedsniveauer, således at alle færdighedsniveauer dækkes (god targeting) og der er lav måleusikkerhed på alle færdighedsniveauer (*standard error of measurement (SEM)*) (se også *anbefaling nr. 45* og *anbefaling nr. 50*).
- Skalaerne/testene bør desuden være kriterievalide, således at der findes de forventede sammenhænge med andre relevante mål.

I tillæg til ovennævnte almene psykometriske krav til pædagogiske måleskalaer, som kan undersøges og sikres med en række forskellige psykometriske metoder og målingsmodeller, skal der også gøres opmærksom på, at tilpasning af opgaverne til en Rasch-model desuden vil betyde, at opgavernes sværhedsgrader vil kunne rangordnes på samme måde for alle elever (psykometrisk krav om homogenitet). Dermed vil man undgå paradokser, hvor dygtigere elever har lavere sandsynlighed for at besvare sværhedsgradsmæssigt lette opgaver korrekt end mindre dygtige elever, og omvendt. Med andre målingsmodeller kan man tilnærme sig, men ikke undgå dette.

Hvis det psykometriske krav om unidimensionalitet ikke kan imødekommes på et testområde, fx læsning, på grund af forskellige underdimensioner, skal hver skala, der måler en underdimension, leve op til ovenstående krav, da det ellers vil give problemer med begrebsvaliditeten i testen.

Hvis kravet om lokal uafhængighed ikke kan opfyldes (dette ses ofte ved testlet-opgaver) anvendes relevante statistiske metoder til at håndtere dette.

Hvis kravet om ingen *differentiel item funktion* ikke kan efterkommes, er det væsentligt, at der tages hånd om dette statistisk, således at der opnås sammenlignelighed af scores for undergrupper. Det skal sikres, at anvendelser til monitorerings- og forskningsformål kan levere korrekte resultater.

Hvis Rasch-modellen vælges som psykometrisk målingsmodel, som i de tidligere nationale test i folkeskolen, kan der desuden opnås den fordel, at sumscoren er *sufficient* for de estimerede Rasch-scores. Det vil således være muligt at anvende disse i tilbagemeldinger til elever. Sumscores har den fordel, at de er lettere at forstå i forhold til testens indhold end Rasch-scores på en logit-skala.

¹⁴ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 9, 15, 17, 46* og *positionsrapport fra Rolf Vegar Olsen med flere*.

Argumenter for at følge

- I. De psykometriske kvalitetskrav for Folkeskolens Nationale Færdighedstest styrkes, hvilket vil påvirke kvaliteten af testresultaterne, kvaliteten af monitoreringsresultater og kvaliteten af forskningsresultater, som baseres på data fra Folkeskolens Nationale Færdighedstest.

Argumenter for ikke at følge

- I. Hvis der gives tilbagemeldinger som sumscores, vil det være nærliggende at tage gennemsnit eller at sammenligne forskelle mellem sumscores. Det vil føre til forkerte konklusioner.

ANBEFALING 11

Det anbefales, at der udvikles en tydelig specifikation af krav til de opgavekommissioner og andre aktører, der involveres i rammesætning, udvikling og afprøvning af opgaver og samlede test til Folkeskolens Nationale Færdighedstest. Der bør være krav om både didaktiske, psykometriske, evalueringsmæssige og faglige kompetencer hos de involverede opgavekommissioner og øvrige involverede aktører for at kunne rammesætte, udvikle og afprøve opgaver og test.

I forbindelse med udvikling, afprøvning og rammesætning af de nye opgaver og de nye test til Folkeskolens Nationale Færdighedstest, bør det tydeligt defineres, at det miljø, som tildeles opgaven, skal have relevante psykometriske, skolefaglige og faglige kompetencer inden for det område/de færdigheder, testen dækker.

Argumenter for at følge

- I. En tydelig kravsspecifikation vil øge muligheden for, at testene bliver relevante for lærere og elever.
- II. En tydelig kravsspecifikation vil bidrage til, at testkvaliteten optimeres.
- III. Et miljø, der integrerer alle relevante kompetencer i det løbende arbejde, er en forudsætning for at udvikle testformater og opgaver, der lever op til høje kvalitetskrav til måleegenskaber og feedback.

ANBEFALING 12

Det anbefales, at der bør være ekstern kvalitetssikring af de nye Folkeskolens Nationale Færdighedstest og dokumentationen af testene. Det anbefales samtidig, at dokumentationen gøres offentligt tilgængelig.

I forbindelse med udviklingen af de nye Folkeskolens Nationale Færdighedstest, uanset om opgaven sendes i udbud, tildeles til udviklere eller holdes internt i Børne- og undervisningsministeriet, bør der også udnævnes/ansættes en eller flere kvalitetssikrere (muligvis en gruppe med forskellige kompetencer), som skal gennemgå dokumentationen af alle trin i testudviklingen ud fra fastlagte kvalitetskriterier. I en første fase kan denne kvalitetssikring lægge vægt på formative aspekter i forhold til forbedringsområder. I en senere fase kan der lægges vægt på tilstrækkeligheden af den psykometriske afprøvning og dokumentation. Dokumentationen bør på sigt offentliggøres (med hemmeligholdelse af eventuelle sensitive elementer i teksten).

Argumenter for at følge

- I. Det er normalt et krav i alle andre sammenhænge, at dokumentation af en tests kvalitet er offentligt tilgængelig.
- II. I tillæg til teknisk dokumentation, som vil kræve specifikke kompetencer at forstå, bør der også fremstilles tilgængelig og forenklet tekst skrevet i et klart sprog, som brugerne af testresultaterne vil kunne forstå.

ANBEFALING 13

Det anbefales, at der gives offentlig adgang til dokumentation af alle væsentlige aspekter af testudviklingen af Folkeskolens Nationale Færdighedstest.

For at sikre gennemsigtighed og mulighed for, at valgene kan efterprøves og diskuteres af forskere og andre med interesse for området, anbefales det, at alle aspekter af testudviklingen og gennemførelsen af Folkeskolens Nationale Færdighedstest, dokumenteres efter internationale standarder (fx *Standards for educational and psychological testing*¹⁵) og stilles til rådighed for download på Børne- og Undervisningsministeriets hjemmeside.

Argumenter for at følge

- I. Udvikling af test er en meget omfattende og kompliceret proces, hvor der tages mange beslutninger, som har konsekvenser for kvaliteten og udformningen af det endelige resultat.

3.4. Testafholdelsen

ANBEFALING 14

Det anbefales, at testvarigheden fastsættes ud fra empiriske afprøvninger, idet det forventes, at det er nødvendigt at tilpasse testtiden til elevernes klassetrin.

Det er et formål, at testene ikke skal tage tid fra undervisning og anden primær aktivitet i skolerne. Hvor lang tid en test tager, hænger sammen med en lang række faktorer, såsom opgaveformat, færdighedsområde, kompleksitet, længde af stimulusmateriale, opgavernes sværhedsgrader og ikke mindst af børnenes alder, idet yngre børn vil have et kortere opmærksomhedsspan og vil udtrættes hurtigere. Testvarigheden er derfor en kompliceret størrelse at få hold på. Det er i høj grad et empirisk spørgsmål, som bør undersøges, og det er derfor nødvendigt at lave nogle empiriske undersøgelser af, hvor lang tid den enkelte test skal vare.

Testens maksimale længde bør være forskellig på forskellige klassetrin – hvor ældre elever vil kunne have længere testtider end yngre elever. En antagelse kunne fx være, at testsessioner for de mindste børn maksimalt kan være 20 minutter, mens de ældste eleverne kan have testsessioner med varighed på op til halvanden time, da udtrætningsgrænsen er forskellig fra de yngste til de ældste. Der vil sandsynligvis være behov for, at test til de yngste elever kan gennemføres gennem flere sessioner med pauser i mellem.

Det bør derfor undersøges empirisk hvor lang testtid, der giver mening på de forskellige klassetrin i de forskellige færdighedstest, da disse vil bestå af forskellige opgavetyper. Ekspertgruppen foreslår, at der afprøves forskellige testtidsgrænser inden for de ovennævnte tidsintervaller med forskellige færdighedsopgavetyper, således at disse kan fastsættes i overensstemmelse med, hvad børnene kan klare med koncentration bevaret, og dermed også om der på nogle klassetrin/i nogle færdighedstest er behov for at dele op i flere testsessioner.

Argumenter for at følge

- I. Hvis test bliver for lange, kan det true testens validitet. Særlig yngre børn vil typisk ikke være i stand til at koncentrere sig i lang tid ad gangen.

¹⁵ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

- II. Hvis test bliver for korte, kan også det true testens validitet, fordi der ikke opnås nok information.
- III. Hensynet til elevernes trivsel.

ANBEFALING 15

Det anbefales generelt, at tiden for den totale testgennemførelse i Folkeskolens Nationale Færdighedstest sættes således, at det store flertal af elever kan nå at blive færdige med alle opgaver i testen, medmindre den færdighed, som skal måles, er afhængigt af, at eleverne bliver presset tidsmæssigt.

Ønsker man eksempelvis at måle elevernes evne til at forstå, tolke og reflektere over en tekst (læseforståelse), så indgår læsehastighed sædvanligvis ikke som en relevant del af færdigheden. I testkonstruktionen bør der derfor tages højde for, at eleverne ikke er færdige på samme tid. Der er dog øvre grænser for den tid, der kan anvendes til at gennemføre testen. De problemer, der kan opstå ved, at nogle elever ikke løser alle opgaver, kan dog løses statistisk.

Argumenter for at følge

- I. Ved at fjerne tidspreset får man som regel en mere valid og præcis måling af det, man ønsker at måle.
- II. Tidspres eller tidsbegrænsning kan medvirke til, at det som måles, ikke er unidimensionelt, da der både måles færdighed og hurtighed, og altså en slags færdighedsmæssig effektivitet.

Argumenter for ikke at følge

- I. Ved at de fleste elever har nok tid til at løse alle opgaver, vil der i de fleste klasser også være et betydeligt antal elever, som er tidligt færdige med testen. Det er derfor vigtigt, at lærerne har pædagogisk tilrettelagte aktiviteter til disse elever.
- II. Det vil give meget uro, og give elever en oplevelse af at være dårlige, hvis de sidder tilbage i lang tid. Derfor vil de svare ukoncentreret og hurtigt.

ANBEFALING 16

Det anbefales, at eleverne gives lov til at bruge alle hjælpemidler, der ikke direkte kompromitterer målingen.

På samme måde som alle hjælpemidler er til rådighed ved flere af folkeskolens skriftlige prøver, så kan de være til rådighed for eleverne ved Folkeskolens Nationale Færdighedstest. Det gælder fx oplæsning, skrivestøtte, lommeregner, plastikmønter, centicubes, ordbøger, internettet (ikke til kommunikation). Hjælpemidler, som er relateret til testens målte indhold (regnemaskine hvis der skal regnes i hovedet, oplæsning hvis der skal måles afkodning), skal ikke være tilladte.

Argumenter for at følge

- I. En test er en kunstig situation, der har til formål at måle, hvad eleverne er i stand til i situationer også uden for testsituationen. I sådanne situationer vil man have adgang til hjælpemidler af forskellig slags – oplæsning, lommeregner, skrivestøtte, ordbøger osv.
- II. Testen skal give læreren et indblik i, hvad eleverne kan også uden for testsituationen, således at undervisningen kan tilrettelægges derefter.

ANBEFALING 17

Det anbefales, at Børne- og Undervisningsministeriet præciserer og uddyber de eksisterende regler¹⁶, vejledning¹⁷ og kriterier for tildeling af særlige vilkår i testsituationen for elever med funktionsnedsættelser og tilsvarende vanskeligheder m.v. Der bør i den forbindelse udarbejdes et særskilt vejledningskatalog, der angiver eksempler på understøttende tiltag og hjælpemidler, der kan anvendes til at kompensere for de konkrete funktionsnedsættelser og tilsvarende vanskeligheder i testsituationen.

Børne- og Undervisningsministeriets eksisterende *Vejledning om særlige prøvevilkår og fritagelse fra folkeskolens prøver*¹⁸ kan anvendes som inspiration for uddybning, præcisering og udarbejdelse af vejledningskatalog. Som minimum bør elever tilbydes samme mulighed for særlige vilkår til Folkeskolens Nationale Færdighedstest, som ved folkeskolens prøver, da Folkeskolens Nationale Færdighedstest har betydning for elevernes videre læreprocesser, skolegang og trivsel.

Særlige vilkår ved Folkeskolens Nationale Færdighedstest bør, som det er tillældet i de gældende regler, udløses ved funktionsnedsættelser og tilsvarende vanskeligheder m.v. Beslutningen om særlige vilkår bør ikke overlades til den enkelte skoleleder alene, men bør træffes i dialog mellem elev, lærer, forældre, skolens leder og eventuelt med inddragelse af pædagogisk-psykologisk rådgivning (PPR), hvis der vurderes konkret behov for det.

Argumenter for at følge

- I. Vil for nogle elever sikre overholdelse af *Lov om forbud mod forskelsbehandling på grund af handicap*. For elever, hvor funktionsnedsættelsen ikke klassificeres som handicap, synes det relevant at give særlige vilkår på grund af funktionsnedsættelsen i henhold til Børnekonventionen og Serviceloven, således at det sikres, at alle elever stilles lige i test gennemførelsen.

ANBEFALING 18

Det anbefales, at der i forbindelse med Folkeskolens Nationale Færdighedstest i læsning imødekommes udfordringerne hos elever diagnosticeret med dysleksi (svarende til rød kategori i Den Nationale Ordblindetest) ved, at de kan fritages fra den del, der tester afkodning.

Den Nationale Ordblindetest tester elevens evne til afkodning, og scorer en elev rød kategori, er det konstateret, at denne elev er stærkt udfordret af afkodningsmæssigt. Det er derfor ikke meningsfuldt at teste denne elevgruppe kontinuerligt i færdigheder, de forventeligt har yderst svært ved at mestre. Dog bør man være opmærksom på, at det på trods af en stærkt øget udfordring med afkodning kan være muligt for elever i rød kategori at øge færdigheder på dette område en lille smule, hvilket for dem kan være ønskværdigt at måle i en test.

Argumenter for at følge

- I. Det er ikke meningsfuldt at teste en gruppe elever i en færdighed, de ud fra en diagnose ikke mestrer, da resultatet ikke giver nye informationer.

¹⁶ Reglerne om særlige vilkår i testsituationen fremgår af §§ 8 – 9 i [Bekendtgørelse nr. 1223 af 31. august 2022 \(Testbekendtgørelsen\)](#)

¹⁷ Vejledning om særlige vilkår i forbindelse med test fremgår i bilaget til *Vejledning om Folkeskolens Nationale Overgangstest – til lærere i dansk og matematik* på Børne- og Undervisningsministeriets hjemmeside: [Folkeskolens Nationale Overgangstest | Børne- og Undervisningsministeriet \(uvm.dk\)](#)

¹⁸ Børne- og Undervisningsministeriet. (2020). *Særlige prøvevilkår og fritagelse fra folkeskolens prøver*. Børne- og Undervisningsministeriet: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf20/dec/210311-vejledning-om-proeve-paa-saerlige-vilkaar.pdf>

- II. At udsætte en gruppe elever for Folkeskolens Nationale Færdighedstest med krav, vi på forhånd ved, de ikke kan honorere, kan påvirke deres selvværd i negativ retning og påvirke deres motivation i forhold til resten af Folkeskolens Nationale Færdighedstest i læsning.

4 Et flertal i ekspertgruppen anbefaler

4.1. Fagligt indhold

ANBEFALING 19

Det anbefales at begrænse indholdet i Folkeskolens Nationale Færdighedstest i matematik til kompetenceområdet *tal og algebra* på 2. klassetrin.

Det anbefales at begrænse indholdet af Folkeskolens Nationale Færdighedstest i matematik på 2. klassetrin til kompetenceområdet *tal og algebra*, men i øvrigt at teste området bredt, som foreslået i *anbefaling nr. 1*. Begrundelserne er, at det er nødvendigt at begrænse testtiden på 2. klassetrin, da eleverne ikke kan koncentrere sig i lang tid ad gangen, og at dette kompetenceområde er centralt for elevernes videre faglige udvikling i matematik.

Argumenter for at følge

- I. I Folkeskolens Nationale Overgangstest er der foretaget tilsvarende begrænsning med samme argumentation. Erfaringerne herfra har hidtil været gode.
- II. På de yngste klassetrin er der stort fokus på kompetenceområdet *tal og algebra* blandt andet fordi, at dette område er afgørende for elevernes videre faglige udvikling i matematik (Nguyen et al., 2016¹⁹; Siegler et al., 2012²⁰).
- III. Området *tal og algebra* udpeges i en nylig rapport fra en ekspertgruppe i matematik (Skott m.fl., 2022²¹), som en ud af otte centrale og aktuelle udfordringer ved matematikfaget i Danmark, der bør handles på. Rapporten bygger blandt andet på TIMSS 2019²², der peger på et markant fald i fagligt niveau blandt de 80 procent dygtigste 4. klasseelever siden den seneste måling i 2015, og at dette fald er større og mere bekymrende, når der ses isoleret på området tal, hvor der er sket et fald på 23 procentpoint siden 2015. Også de 20 procent svageste elever oplever større udfordringer i området tal, end de gør i de to andre områder måling og geometri (Kjeldsen, Kristensen & Christensen, 2020²³).

¹⁹ Nguyen, Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). *Which preschool mathematics competencies are most predictive of fifth grade achievement?* Early Childhood Research Quarterly, 36 (3), s. 550–560.

²⁰ Siegler, Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). *Early Predictors of High School Mathematics Achievement*. Grantee Submission, 23 (7), s. 691–697.

²¹ Skott, C. K., Christensen, B. K., Zacho, L., Thompsen, M., Teglskov, R., Øhrstrøm, A., Rønne, M., Riisgaard, T. B., Tange, B., Granerud, L. S., Hørup, L. A., Winsløw, C., Sunde, P. B., & Brockhoff, P. (2022). *Fælles udvikling af matematik: Om udfordringer og mulige løsninger i forhold til matematik i grundskolen, på de gymnasiale uddannelser og på erhvervsuddannelserne*. Børne- og Undervisningsministeriet.

²² Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center: <https://timss-andpirls.bc.edu/timss2019/international-results/>

²³ Kjeldsen, Kristensen & Christensen (2020). *Matematik og natur/teknologi i 4. klasse. Resultater af TIMSS-undersøgelsen 2019*. Aarhus Universitetsforlag: [Matematik og natur/teknologi i 4. klasse \(unipress.dk\)](https://matematik.aaun.dk/)

ANBEFALING 20

Det anbefales, at læsefærdigheden afkodning i 2., 3., 4. og 6. klasse testes med et særskilt, obligatorisk opgaveformat, der både måler præcision og hastighed.

Afkodning, dvs. genkendelse af ordbetydning fra bogstaver uafhængigt af konteksten, er fundamentet for tekstlæsning i dansk og alle andre fag og har derfor særstatus. Selvom mange børn bliver rimeligt præcise til at afkode i løbet af nogle år, så kan det tage lang tid at opøve tilstrækkeligt hurtig afkodning. Afkodningen skal være hurtig for at blive et nyttigt og lettilgængeligt redskab i læsning af tekster. Fx angiver Fælles Mål hurtig og sikker læsning som mål, og en læsehastighed på 120 ord/minut som opmærksomhedspunkt for slutningen af 4. klasse.

Argumenter for at følge

- I. Det er nødvendigt at adskille afkodning fra bredere læseforståelse. Fx har elever med andet modersmål end dansk gennemsnitlig afkodningsudvikling, men kan have vanskeligheder med læseforståelse på grund af begrænsninger i dansk sprogfærdighed. Omvendt kan elever med svage afkodningsfærdigheder, fx nybegynderlæsere og ordblinde, have udmærket forståelse af oplæste tekster. Disse grupper af elever har behov for forskellige pædagogiske tiltag.
- II. Hvis afkodning testes selvstændigt, kan øvrige resultater i læsning gives en pædagogisk anvendelig tolkning. Fx vil et svagt resultat i delfærdigheden tekstforståelse sammenholdt med et gennemsnitligt afkodningsresultat indikere, at sprogforståelsesvanskeligheder er en mulig årsag til læsevanskelighederne.
- III. Hvis afkodning testes separat, så kan øvrig læsning testes med mindre fokus på hastighed.
- IV. Hvis hastighed inddrages, så er det muligt også at måle højere færdighedsniveauer i afkodning uden at bruge ord, der er ukendte for klassetrinnet. Man kan således undgå nogle velkendte problemer fra de adaptive nationale test.
- V. Afkodning kan sandsynligvis testes rimeligt pålideligt i løbet af kort tid.

Argumenter for ikke at følge

- I. Det er en teknisk ekstraopgave at tilføje opgaveformater med tid.
- II. Direkte undervisning i afkodning fylder uforholdsmæssigt meget i megen danskundervisning i dag (Bundsgaard, Buch & Fougts 2017²⁴). Nogle elever har glæde af det, men for mange er det unødvendigt og fører til oplevelse af kedsomhed. Hvis Folkeskolens Nationale Færdighedstest tester afkodning, risikeres det, at lærerne bruger endnu mere tid på at terpe afkodning til skade for undervisning i alle de andre, vanskeligere, læsefærdigheder.
- III. Beslutning om afgrænsning af, hvad der skal testes, bør tages af dem, der udvikler testen, således at beslutningen kan ske på baggrund af grundige, faglige overvejelser, som udfoldes i et rammeværk for testen, og med afvejning i forhold til de øvrige rammer for testen.

ANBEFALING 21

Det anbefales, at læsning testes som praksisrettet færdighed.

Folkeskolens Nationale Færdighedstest skal, ligesom Fælles Mål og anerkendte internationale læsetest, betragte læsning som en praksisrettet færdighed med fokus både på at finde tekster, forberede læsning, læse med mange strategier (uddrage informationer, sammenholde informationer, skimme, skabe overblik, fokusere, læse i relation til andre modaliteter mv.), forholde teksten til kommunikationssituationen og sætte teksten i sammenhæng.

²⁴ Bundsgaard, J., Buch, B., & Fougts, S. S. (2017). *De anvendte læremidlers danskfag belyst kvantitativt*. I J. Bremholm, J. Bundsgaard, S. Skov Fougts, & A. Karlskov Skyggebjerg (red.), *Læremidlernes danskfag*, s. 28-54. Aarhus Universitetsforlag. Didaktiske studier, Bind 1.

Argumenter for at følge

- I. Læsning er en færdighed, som udfoldes i alle livets forhold. Udøvelsen af færdigheden hænger tæt sammen med den sammenhæng, den bruges i. Læser man tekster i naturfag, skal man forbinde de verbalsproglige ytringer med diagrammer, modeller og illustrationer for at forstå indholdet. Læser man tekster i samfundsfag, skal man kunne sammenholde indholdet i tabeller og oversigter med den verbalsproglige sammenfatning. Læser man tekster i madkundskab eller i håndværk og design, skal man kunne sammenholde korte, komplekse verbalsproglige tekster integreret med diagrammer og tegninger. Læser man tekster for at finde frem til en konkret information, skal man kunne overskue, skimme, nærlæse og vurdere kildens anvendelighed. Og man skal kunne finde frem til teksten på biblioteket eller på internettet. Læser man tekster fra kommercielle eller ideologiske afsendere, skal man kunne vurdere interesserne bag teksten, og sammenholde den med sine egne interesser. Læser man argumenterende tekster i sociale medier eller i aviser, skal man kunne gennemskue og vurdere argumentationen. Læsning er med andre ord en færdighed, der opøves gennem mødet med tekster i mange genrer og i mange situationer. Derfor er det vigtigt, at færdigheden læsning testes i virkelighedsnære og varierede praksisser.
- II. Læsning opfattes i Fælles Mål som en kompleks færdighed bestående af færdighedsområderne *Finde tekst, Forberedelse, Afkodning, Sprogforståelse, Tekstforståelse, Sammenhæng*. Der er tillige beskrevet læsefærdigheder under *Fortolkning og Kommunikation*.
- III. Internationalt har læsetest i dag fokus på at teste færdigheder i stil med de her beskrevne. Fx lyder PISAs definition af læsning: At være i besiddelse af læsekompetence (*reading literacy*) vil sige, at man kan forstå, bruge, vurdere, reflektere over og engagere sig i tekster, så man kan opnå sine mål, udvikle sin viden og sine muligheder og deltage aktivt i samfundslivet (OECD, 2019, Vol. I, kap. 1²⁵). Og PIRLS' definition lyder "Læsekompetence er kompetencen til at forstå og anvende de skriftsproglige udtryksformer, der kræves af samfundet og/eller værdsættes af den enkelte person. Læsere kan danne sig en forståelse fra vidt forskellige tekster. De læser for at lære, for at deltage i fællesskaber med andre, der læser, hvad enten det er i skolen, i fritiden eller for deres egen fornøjelses skyld." (Mullis & Martin, 2015²⁶).
- IV. Hvis testen kun har fokus på de tekniske aspekter af læsning i ren verbalsproglig form, vil lærerne rette deres undervisning mod læsning uafhængig af sammenhængen. Det vil betyde, at det er op til eleverne selv at udvikle færdigheder i at sammenholde verbalsproglig tekst med alle de andre meddelelsesformer teksten indgår i sammenhæng med. Det vil gå mest ud over de svageste elever, som ikke har forældre til at hjælpe sig. Det er derfor vigtigt, at læsning testes som praksisrettet færdighed.

Argumenter for ikke at følge

- I. Beslutning om afgrænsning af, hvad der skal testes, bør tages af dem, der udvikler testen, således at beslutningen kan ske på baggrund af grundige, faglige overvejelser, som udfoldes i et rammeværk for testen, og med afvejning i forhold til de øvrige rammer for testen.
- II. Hvis læsning testes som praksisrettet færdighed i forskellige sammenhænge, vil der komme mange dimensioner med. Det kan medføre øgede krav til testen og tvivl om, hvorvidt de samme dimensioner måles fra år til år.
- III. Det er ikke adækvat beskrevet, hvad der menes med praksisrettet færdighed, og det giver umiddelbart ikke mening at ville inkludere eksempelvis "Læser man tekster fra kommercielle eller ideologiske afsendere, skal man kunne vurdere interesserne bag teksten, og sammenholde den med sine egne interesser. Læser man argumenterende tekster i sociale medier eller

²⁵ OECD. (2019). *PISA 2018 Results (Volume I). What Students Know and Can Do*. OECD.
<https://doi.org/10.1787/5f07c754-en>

²⁶ Mullis, I. V. S., & Martin, M. O. (Red.). (2015). *PIRLS 2016 assessment framework second edition (2nd edition)*. TIMSS & PIRLS International Study Center.

i aviser, skal man kunne gennemskue og vurdere argumentationen”, da vurdering ikke er en del af læsning, sammenholdt med egne interesser ikke er en del af læsning, at kunne gennemskue ikke er en del af læsning osv. Med anbefalingen foreslås det tilsyneladende at teste en blanding af flere begreber på en gang og ikke en eller flere afgrænsede færdigheder, som kan testes med en unidimensionel skala/test (jf. *anbefaling nr. 10* og *anbefaling nr. 15*).

ANBEFALING 22

Det anbefales, at opgaveformater og delopgaver i Folkeskolens Nationale Færdighedstest i læsning målrettes pålidelig feedback i færdighedsområderne *afkodning, sprogforståelse og tekstforståelse* i Fælles Mål for danskfaget.

Eftersom Folkeskolens Nationale Færdighedstest ikke omhandler hele danskfaget, kan opgaverne målrettes tilbage om et mindre antal generelle færdighedsområder, der specifikt handler om læsning i Fælles Mål, primært *afkodning, sprogforståelse og tekstforståelse*, og som kan genfindes i internationale læserammeværk som PISA og PIRLS. Ved at målrette opgaver mod et begrænset antal delfærdigheder, kan man give mere pålidelig tilbagemelding om disse færdigheder. Men opgaverne bør indlejres i et bredt udsnit af klassetrinrelevante tekstgenrer (fx informerende tekster, skønlitteratur- og websider med kommentarspor) og læseaktiviteter med inspiration fra flere af færdighedsområderne i læsning (fx navigation rundt i tekster), så der opmuntres til bredde i læseaktiviteter i undervisningen. Opgaverne kan altså trække på et bredt udsnit af færdigheder, uden at der gives tilbagemelding om alle disse færdigheder.

I det omfang opgavetyper og delopgaver skal danne grund for delskalaer eller færdighedsbeskrivelser, bør opgaverne udvikles specifikt til at understøtte sådanne delskalaer. Dette er i modsætning til, at opgavernes relatering til delskalaer/færdigheder tilføjes efterfølgende.

Argumenter for at følge

- I. Jo flere opgaver, der designes specifikt til at teste en færdighed, jo mere pålideligt kan man tilbagemelde om færdigheden.

Argumenter for ikke at følge

- I. Stor variation i opgaveformater kan være en særlig udfordring for svage elever, hvis de hele tiden på ny skal finde ud af, hvad de skal. Variation bør afvejes mod tilgængelighed.
- II. Ved at kræve, at alle disse tre delfærdigheder skal være inkluderet i testen, indfører man de facto, at der skal være tre test inden for området læsning. Dette må derfor afvejes mod hensynet til, at den totale testtid er begrænset.
- III. Et alternativ er, at lærere anvender andre test til evaluering og måling af afkodning og sprogforståelse, og at Folkeskolens Nationale Færdighedstest i læsning dermed alene prioriterer evaluering og måling af læseforståelse.
- IV. En test som alene måler sprogforståelse og tekstforståelse, vil medføre, at undervisningen i endnu højere grad end i dag, har fokus på tekniske aspekter af læsning, hvilket betyder, at eleverne ikke i tilstrækkeligt omfang undervises i læsning som praksisrettet færdighed.
- V. Læsning i skolens andre fag kræver færdigheder i at læse med forskellige strategier, vælge sin læsestrategi, læse multimodalt, sætte det læste i sammenhæng, forholde sig kildekritisk, gennemskue teksters indlejrede budskaber osv. Det er vigtigt, at læseundervisningen i dansk forbereder eleverne på dette. Og derfor er det vigtigt, at testen tester dette.
- VI. Både internationalt og i Danmark ser man ofte elever, som tilsyneladende læser fint, men når de når omkring 4. klasse, viser det sig, at de ikke har udviklet læsefærdigheder til at læse tekster i de andre fag. Det omtales ofte som *the fourth-grade slump*, hvor man går fra at "lære at læse" til at "læse for at lære".

- VII. Beslutning om afgrænsning af, hvad der skal testes, bør tages af dem, der udvikler testen, således at beslutningen kan ske på baggrund af grundige, faglige overvejelser, som udfoldes i et rammeværk for testen, og med afvejning i forhold til de øvrige rammer for testen.

ANBEFALING 23

Det anbefales, at den frivillige test i dansk som andetsprog indeholder opgaver, som tester elevens sprog meget tæt på alderssvarende elevers hverdags- og fagsprog.

For at kunne afdække tosprogede elevers kompetencer og potentialer i forhold til at tilegne sig dansk, er det vigtigt, at testens opgaver, som tester sprog og sprogbrug, er så tæt på både det anvendte hverdags- og fagsprog, der gør sig gældende på det pågældende klassetrin, som muligt.

ANBEFALING 24

Det anbefales, at hovedparten af opgaverne i Folkeskolens Nationale Færdighedstest genanvendes, men at et lille udsnit udskiftes fra år til år.

Den pædagogiske brug af test forudsætter, at lærere har et indgående kendskab til opgaverne og til, hvad de tester. Derved kan læreren ved blot at iagttage, hvilke opgaver en elev har besvaret rigtigt og forkert, hurtigt identificere hvilke særlige udfordringer den enkelte elev har. Enkelte opgaver udskiftes for, at testen følger med tiden, og således at forsøg på at forberede eleverne på testen, kan afsløres.

Argumenter for at følge

- I. Når der kun skal udvikles et mindre sæt opgaver fra år til år, kan der lægges en fokuseret indsats i at udvikle gode, teknisk komplekse, grafisk gennemtænkte og velafprøvede opgaver.
- II. Lærere kan løbende udveksle erfaringer med kolleger om de faglige delelementer, de forskellige opgaver tester og diskutere, hvordan undervisningen i disse delelementer kan foregå.
- III. Den faglige dialog om Folkeskolens Nationale Færdighedstest kan bygge på et velkendt og fælles grundlag.
- IV. En velkendt test vil understøtte lærernes pædagogiske arbejde og modvirke oplevelsen af, at testen er til for kontrolformål.
- V. Hvis de nye opgaver er hemmelige frem til testen, kan lærere eller elever, der har "snydt" ved at forberede sig på den konkrete test, afsløres (de vil klare sig uforholdsmæssigt dårligere på de nye opgaver).

Argumenter for ikke at følge

- I. Denne anbefaling hviler på en stærk antagelse om, at alle kommuner, skoler eller lærere tænker, at elevresultaterne fra Folkeskolens Nationale Færdighedstest ikke har store konsekvenser for den enkelte elev, lærer eller skole. Erfaringer tilsiger, at dette ikke er en realistisk antagelse.
- II. Usikkerheden i resultatet for en enkeltelev på en enkeltopgave er for stor til, at "læreren ved blot at iagttage hvilke opgaver en elev har besvaret rigtigt og forkert, hurtigt kan identificere hvilke særlige udfordringer den enkelte elev har".
- III. Ved at opgavetyper genanvendes, men indholdet i den enkelte opgave ikke kendes af lærere og elever på forhånd, kan man tilgodese hensynet til, at læreren kan bruge opgavetyper til at vurdere, hvor elevgruppen eventuelt har udfordringer – samtidig med, at formålet om, at kommuner og skoler kan bruge resultaterne fra testene som grundlag for kvalitetsudvikling bedre tilgodeses.
- IV. Anbefalingen lægger op til *teaching to the test*.
- V. Med korte færdighedstest er det ikke realistisk at genbruge hele test, da hukommelseffekten vil være for stor.
- VI. Anbefalingen forhindrer monitorering, frigjort fra *teaching to the test*.

ANBEFALING 25

Det anbefales, at opgaver i Folkeskolens Nationale Færdighedstest fornyes og udskiftes jævnligt.

Da opgaverne i Folkeskolens Nationale Færdighedstest er tilgængelige, kan der være en risiko for, at eleverne bliver undervist efter netop de opgaver/spørgsmål, der stilles, hvis der ikke jævnligt sker en fornyelse og udskiftning af opgaverne. Hvis der med jævne mellemrum sker udskiftning af opgaver/spørgsmål, kan det måske minimere denne form for *teaching to the test*. Desuden kan opgaver/spørgsmål på den måde udvikle sig sammen med tiden.

Den jævnlige udskiftning af opgaver/spørgsmål bør ske uden, at der samtidig sker store udskiftninger i selve opgavetyperne, således at eleverne oplever en genkendelighed i opgavetyperne.

Argumenter for at følge

- I. Det sikres, at opgaverne og det faglige indhold i Folkeskolens Nationale Færdighedstest følger med udviklingen og tiden.
- II. Jævnlig udskiftning af opgaver har samtidig til formål at forsøge at minimere *teaching to the test*.

4.2. Testresultaternes anvendelse

ANBEFALING 26

Det anbefales, at de bestemmelser vedrørende fortrolighed om testresultater, som er nævnt i folkeskoleloven, også fastholdes fremover for Folkeskolens Nationale Færdighedstest.

Testresultater for den enkelte elev, grupper af elever, hold, klasser, skoler, kommuner og regioner m.v. er fortrolige. Undtaget er opgørelser på landsplan og de informationer om den enkeltes testresultater, som tilflyder den enkelte elev og dennes forældre. For de nærmere bestemmelser henvises til folkeskolelovens § 13 og § 55 b.

Argumenter for at følge

- I. Ved at undlade en rangordning af skoler, kommuner eller regioner fra statslig side signaleres prioritering af Folkeskolens Nationale Færdighedstest som pædagogiske værktøjer, og lærernes pædagogiske anvendelse af resultaterne med den enkelte elev værdsættes og understøttes.
- II. Det minimerer den offentlige nedgørelse af skoler og kommuner i medier, og modvirker tillige *teaching to the test*, idet skolernes rangordning ikke er kendt.
- III. I Norge er det muligt for pressen at generere ranglister over skoler og skrive om disse. Der ses en udvikling væk fra dette, som opleves positivt.

ANBEFALING 27

Det anbefales, at der sikres validitet i alle anvendelser af Folkeskolens Nationale Færdighedstest.

Alle anvendelser af Folkeskolens Nationale Færdighedstest (som initieres eller sanktioneres af ministerier, kommuner, forskningsinstitutioner eller andre offentlige institutioner) skal være valide. Det vil sige, at udsagn om skoler eller indsatsers succes skal relateres tydeligt i forhold til det indhold, der er testet. Når autoriteter (som fx offentlige institutioner) anvender resultater fra Folkeskolens Nationale Færdighedstest, påtager de sig en forpligtelse til at begrunde, hvorfor den givne anvendelse er valid. Testen er udviklet til at måle en given færdighed, så hvis den fx bruges til at sige noget om en skoles effektivitet

eller en lærers undervisning, så skal det begrundes, hvorfor testen har udsigelseskraft i den sammenhæng.

Argumenter for at følge

- I. Folkeskolens Nationale Færdighedstest tester kun et afgrænset område af de meget omfattende faglige mål. Det er derfor vigtigt, at offentligheden ikke får indtryk af, at Folkeskolens Nationale Færdighedstest er udtryk for niveauet af hele den brede faglighed. Det gælder både i forhold til enkeltelever og i forhold til skoler og i forhold til undersøgelser og forskningsresultater. Anvendelse af testresultater til udsagn om succes og forskningsmæssige sammenhænge skal bygge på dyb forståelse af, hvad der er testet og hvordan, således at påstande og konklusioner kan nuanceres og forklares af de, der anvender data.

ANBEFALING 28

Det anbefales, at aktører, der anvender Folkeskolens Nationale Færdighedstest, gør sig klart, at testresultaterne handler om et afgrænset antal centrale færdigheder.

Disse centrale færdigheder udgør kun en afgrænset delmængde af de meget talrige og omfattende faglige mål i folkeskolen og dækker tillige kun en begrænset del af folkeskolens formål.

Det er derfor vigtigt, at beslutningstagere og offentligheden ikke får indtryk af, at testen angiver niveauet af hele den brede faglighed. På den baggrund advares mod udsagn, hvor Folkeskolens Nationale Færdighedstest uden videre bringes til at stå for eksempelvis kvaliteten eller effektiviteten af en lærers undervisning, en skole eller et skolesystem.

Argumenter for at følge

- I. Hvis det lykkes at sikre, at udsagn om Folkeskolens Nationale Færdighedstest afgrænses til at beskrive et afgrænset antal centrale færdigheder, kan dette sammen med andre tiltag være med til at dæmpe effekter i retning af *teaching to the test*.

Argumenter for ikke at følge

- I. Anbefalingen er for bred, da "aktører" kan være hvem helst (forskere, m.v., som har frihed til at anvende resultaterne, som de vil).

ANBEFALING 29

Det anbefales, at Børne- og Undervisningsministeriet udarbejder vejledninger til forskere (og konsulenthuse), statslige og kommunale embedsfolk og skoleledelser om, hvordan data fra Folkeskolens Nationale Færdighedstest kan anvendes og fortolkes.

Børne- og Undervisningsministeriet (eller testudvikleren) skal udarbejde grundige og detaljerede vejledninger i, hvordan data fra Folkeskolens Nationale Færdighedstest kan anvendes og fortolkes.

Vejledningerne skal blandt andet indeholde definitioner af de målte områder, forklaringer på hvad der måles og ikke måles, beskrivelser af hvordan der måles, forklaringer på hvilke statistiske modeller, der typisk vil blive taget i brug og hvad man ikke kan gøre med data, medmindre man har videnskabeligt belæg for andet.

Argumenter for at følge

- I. Vejledningerne kan give grundlag for at påpege misbrug af data fra Folkeskolens Nationale Færdighedstest.

ANBEFALING 30

Det anbefales, at der sikres brugervenlighed i tilgængeligheden af data fra Folkeskolens Nationale Færdighedstest, så de kan trækkes nemt og overskueligt af relevante brugere.

Testresultaterne fra Folkeskolens Nationale Færdighedstest skal bruges til ledelsesinformation på kommunalt og nationalt niveau, så der er mulighed for at følge, hvordan eleverne klarer sig. Der er derfor behov for, at data på skole- og kommuneniveau kan tilgås nemt af relevante brugere, så tidsforbruget i forbindelse med at fremskaffe data til ledelsesinformationen minimeres.

Argumenter for at følge

- I. Det fremgår af den politiske *aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen*, at formålet blandt andet er, at det kommunalt og på landsplan skal være muligt at følge med i, hvordan eleverne klarer sig fagligt. Om Folkeskolens Nationale Færdighedstest står der, at kommunerne skal kunne følge elevernes faglighed på de enkelte skoler, og nationalt skal det faglige niveau kunne følges på skole- og kommuneniveau. Derudover fremgår det af aftalen, at evaluerings- og bedømmelsessystemet skal bidrage til mindre bureaukrati.

Argumenter for ikke at følge

- I. Resultaterne fra Folkeskolens Nationale Færdighedstest er ikke lette at fortolke og anvende til beslutninger. Hvis man får den opfattelse, er det meget sandsynligt, at man kommer til ikke-underbyggede konklusioner. Det bør derfor ikke være let at tilgå og bruge dem.

ANBEFALING 31

Det anbefales, at de samlede testresultater på skole- og kommuneniveau præsenteres i et format, som er let forståeligt.

Testresultaterne fra Folkeskolens Nationale Færdighedstest skal bruges til ledelsesinformation på kommunalt og nationalt niveau, så der er mulighed for at følge, hvordan eleverne klarer sig. Der er derfor behov for, at data på skole- og kommuneniveau fremstår umiddelbart forståeligt, så der ikke er en omfattende oversættelsesopgave, når de skal bruges til ledelsesinformation kommunalt og nationalt.

Argumenter for at følge

- I. Det fremgår af den politiske *aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen*, at formålet blandt andet er, at det kommunalt og på landsplan skal være muligt at følge med i, hvordan eleverne klarer sig fagligt. Om Folkeskolens Nationale Færdighedstest står der, at kommunerne skal kunne følge elevernes faglighed på de enkelte skoler, og nationalt skal det faglige niveau kunne følges på skole- og kommuneniveau. Derudover fremgår det af aftalen, at evaluerings- og bedømmelsessystemet skal bidrage til mindre bureaukrati.
- II. Der forventes ikke at blive tilført midler, og de ressourcer, der bruges til at oversætte testresultater til relevant ledelsesinformation, kan ikke bruges til andet.

Argumenter for ikke at følge

- I. Resultaterne fra Folkeskolens Nationale Færdighedstest er ikke lette at fortolke og anvende til beslutninger. Hvis man får den opfattelse, er det meget sandsynligt, at man kommer til ikke-underbyggede konklusioner. Det bør derfor ikke være let at tilgå og bruge dem.

ANBEFALING 32

Det anbefales, at de samlede testresultater på skole- og kommuneniveau præsenteres på en måde, hvor det tydeligt fremgår, at testresultaterne fra Folkeskolens Nationale Færdighedstest har en begrænset udsigelseskraft.

Folkeskolens Nationale Færdighedstest kommer ikke til at kunne måle elevernes udbytte af alle dele af de pågældende fag. Testresultaterne kan kun sige noget om det, der reelt er målt. Det er derfor afgørende, at det fremgår tydeligt, hvilke dele af faget, testen har udsigelseskraft om – og hvilke dele testene ikke har udsigelseskraft om. Det bliver dermed også anskueliggjort, at testresultaterne fra Folkeskolens Nationale Færdighedstest kun bør indgå som ét parameter i den samlede ledelsesinformation.

Argumenter for at følge

- I. Det fremgår af den politiske *aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen*, at formålet blandt andet er, at det kommunalt og på landsplan skal være muligt at følge med i, hvordan eleverne klarer sig fagligt. Om Folkeskolens Nationale færdighedstest står der, at kommunerne skal kunne følge elevernes faglighed på de enkelte skoler, og nationalt skal det faglige niveau kunne følges på skole- og kommuneniveau.
- II. Resultaterne af de tidligere nationale test har været brugt på en måde, hvor de blev taget til indtægt for folkeskolens kvalitet helt generelt både kommunalt og nationalt (blandt andet i forbindelse med kvalitetsrapporterne og i følgeforskningen for folkeskolereformen).

ANBEFALING 33

Det anbefales, at resultaterne fra Folkeskolens Nationale Færdighedstest ikke bruges alene som mål for udvikling af folkeskolen kommunalt og nationalt. Testresultaterne kan indgå som ét succeskriterium blandt andre, når fx en kommunal læseindsats skal evalueres. Men testresultaterne kan ikke bruges som eneste indikator på, om en indsats har været vellykket eller ej.

Folkeskolens Nationale Færdighedstest skal, jf. *aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen* bidrage til, at kommunerne kan følge elevernes faglighed på de enkelte skoler. Nationalt skal det faglige niveau kunne følges på skole- og kommuneniveau. Men test, feedback, karakterer og andre former for bedømmelse og evaluering er en integreret del af evalueringskulturen i folkeskolen og udgør til sammen et billede af elevernes faglige niveau. Testresultaterne fra Folkeskolens Nationale Færdighedstest udgør kun ét parameter i den forbindelse.

Argumenter for at følge

- I. Det er nemt kommunalt og nationalt at opstille kvantitative måltal. Ved at lave en anbefaling om ikke at bruge kvantitative mål for testresultaterne, bliver *teaching to the test* forhåbentlig minimeret.
- II. Ved at anbefale at testresultaterne alene kan bruges som ét succeskriterie blandt andre, bliver tendensen til at se testresultaterne som den endegyldige sandhed om elevens/skolens/skolevæsnet/folkeskolens faglige niveau forhåbentlig minimeret.

ANBEFALING 34

Det anbefales, at der ikke kan pålægges lokale aktører, hvordan resultaterne fra Folkeskolens Nationale Færdighedstest skal bruges.

Selve testen og resultaterne fra Folkeskolens Nationale Færdighedstest giver mulighed for mange anvendelser. Hvis man fra centralt hold (ministerium, kommune, skoleleder, faglig vejleder) pålægger lokale aktører (kommuner, skoleledere, vejledere, lærere) at anvende disse efter særlige modeller, fratager man de lokale aktører friheden til at udøve deres professionelle skøn og anvende test og resultater på de måder, som vurderes hensigtsmæssige i den lokale kontekst.

Argumenter for at følge

- I. Kommuner, skoler og lærere skal have frihed til at udøve deres professionelle praksis.
- II. Der er politisk ønske om at give skolerne frihed og om at reducere bureaukrati.
- III. Der er meget kort tid til mange aktiviteter i skolen, fx mindre end et kvarter til samtaler med forældre. Hvis der udarbejdes generelle retningslinjer for, hvordan Folkeskolens Nationale Færdighedstest skal fremlægges for forældrene, kan lærerne blive forhindret i at tage ting op, som er af afgørende betydning for eleven og for samarbejdet. Hvis skoleledelsen pålægges bestemte måder at tage resultaterne op med medarbejderne, kan det forhindre, at de indgår på lokalt meningsfulde måder.

Argumenter for ikke at følge

- I. Det bør være muligt på kommune- og skoleniveau at aftale, hvordan Folkeskolens Nationale Færdighedstest skal bruges. Sådanne aftaler vil kunne opleves som et oplæg af enkelte skoler/lærere.
- II. Det fremstår ikke tydeligt, hvad der menes med "lokale aktører og "brug af resultaterne fra Folkeskolens Nationale Færdighedstest", herunder med hvilket formål, og det kan således pege i mange retninger af brug.
- III. Testresultater bør anvendes på de måder, som beskrives i vejledende materiale og ikke på andre skønsmæssige ansatte måder, da det vil give mulighed for uhensigtsmæssigt brug/misbrug.

ANBEFALING 35

Det anbefales, at anvendelsen af testresultater på kommuneniveau er læringsorienteret og dialogbaseret.

Anvendelsen af fag- og skoleaggregerede testresultater på kommuneniveau bør være læringsorienteret og dialogbaseret. Kommunerne bør gives lokal selvbestemmelse med hensyn til tilrettelæggelse af dialogen mellem forvaltningen og skolerne. Testresultater kan sammen med andre typer data med fordel indgå som en del af baggrundsmaterialet for skoleudviklingssamtalerne. Det bør sikres, at der sker en erfaringsformidling på tværs af skoler og kommuner – fx i form af udarbejdelse af inspirationskataloger - vedrørende god praksis i anvendelse af testresultater og andre data.

Argumenter for at følge

- I. En lærings- og dialogbaseret tilgang med lokal selvbestemmelse kan bidrage til lokalt ejerskab og engagement og understøtte kvalitetsudvikling.

Argumenter for ikke at følge

- I. Denne anbefaling betyder, at der pålægges tidsforbrug, som forhindrer, at tiden bruges på det, der anses for hensigtsmæssigt lokalt.

ANBEFALING 36

Det anbefales, at anvendelsen af testresultater på skoleniveau er læringsorienteret og dialogbaseret.

Anvendelsen af testresultater vedrørende de enkelte klassers færdighedsniveau bør på skoleniveau være læringsorienteret og dialogbaseret. Praksis bør være lokalt forankret og kan med fordel tænkes ind i den enkelte skoles rutiner for møder i lærerteams, lærerkollegier og ledelsesteams. Testresultater bør anvendes i kombination med andre typer data.

Argumenter for at følge

- I. En lærings- og dialogbaseret tilgang med lokal selvbestemmelse kan bidrage til lokalt ejerskab og engagement og understøtte kvalitetsudvikling.

Argumenter for ikke at følge

- I. Denne anbefaling betyder, at der pålægges tidsforbrug, som forhindrer, at tiden bruges på det, der anses for hensigtsmæssigt lokalt.

ANBEFALING 37

Det anbefales, at der udarbejdes et vejledningsmateriale til opkvalificering af de faglige vejledere, så de kan støtte lærerne i tolkning, analyse, formidling og pædagogisk anvendelse af resultater fra testene.

Der skal udarbejdes et vejledningsmateriale, fx i form af videoer, webinarer og skriftligt materiale, som kan bruges til opkvalificering af de faglige vejledere i forhold til at afkode testresultaterne og vejlede fagkollegaer i at analysere, fortolke og formidle testresultaterne til elever og forældre. Vejledningsmaterialet skal også angive, hvordan resultaterne kan anvendes til at iværksætte relevante tiltag på både elev- og klasseniveau.

Argumenter for at følge

- I. Opkvalificering af de faglige vejledere kan sikre en systematisk, valid og ensartet efterbehandling af testresultater.
- II. Ved inddragelse af vejledere, sikres en faglig dialog mellem lærer og vejleder med henblik på opfølgning i undervisningen.

Argumenter for ikke at følge

- I. Der bliver et ekstra led mellem testresultatet og lærerens tilbagemelding til elever henholdsvis samtale med forældre. Dette vurderes unødvendigt, da testresultatets form, skriftlige vejledninger og lærernes direkte opkvalificering bør være tilstrækkeligt.

ANBEFALING 38

Det anbefales, at der udarbejdes vejledningsmateriale til lærere, vejledere, skoleledere, forældre og andre brugere om indholdet i Folkeskolens Nationale Færdighedstest.

Brugerne skal støttes i at bruge testens resultater konstruktivt og meningsfuldt på alle niveauer. Vejledningerne skal blandt andet indeholde de generelle intentioner med testen, definitioner af de målte områder, beskrivelser af, hvordan der måles, en overordnet oversigt over testens opgavetyper og indhold, samt de områder, som testen ikke måler. Vejledningerne skal formuleres i et lettilgængeligt sprog eventuelt i videoer, uden at det går ud over fagligheden. Til brugere med særligt ansvar eller interesse,

fx faglige vejledere, bør der være supplerende materiale i form af webinarer og skriftligt materiale. Materialet bør desuden forsynes med henvisninger (links og referencer) til mere tekniske beskrivelser målrettet fagfolk.

I fremhævelse af, hvad testene *ikke* måler, bør også ligge en understregning af, at testene ikke kan bruges som et generelt mål for udvikling i fagene endside skolernes samlede kvalitet.

ANBEFALING 39

Det anbefales, at testresultaterne vedrørende den enkelte elev og for klassen gengives, så det understøtter lærerens pædagogiske arbejde med den enkelte elev samt med klassen bedst muligt og forenkler formidlingen til forældrene.²⁷

Hver enkelt lærer, som skal arbejde pædagogisk formativt med resultaterne fra Folkeskolens Nationale Færdighedstest, skal gøre dette både på enkeltlevniveau og på klasseniveau. Det er derfor nødvendigt, at resultaterne på begge disse niveauer gengives, så det understøtter lærerens pædagogiske arbejde med den enkelte elev og med klassen bedst muligt, dvs. enkelt, intuitivt forståeligt og kortfattet.

Det foreslås derfor, at resultaterne (jf. *anbefaling nr. 51*) på elevniveau gengives grafisk inklusiv tekst i et resultatark, der beskriver elevens udviklingsområder m.v. på en enkelt side. Dette er set i andre testsystemer. Dette resultatark kan anvendes sammen med eleven, og det kan anvendes i samtale med forældrene. Resultatarkene på elevniveau suppleres med et klasseoverblik, hvor resultaterne fra klassens elever plottes i forhold til hinanden, og suppleres med de typiske udviklingsområder på de forskellige score-niveauer. Klasseoversigten bør udelukkende være et redskab for læreren og dermed ikke give anledning til, at elever kan sammenligne sig med hinanden, eller at forældrene kan sammenligne klassens elever. Desuden bør tilbagemeldingsmaterialet suppleres med en generel vejledning til læsning og tolkning af resultater fra Folkeskolens Nationale Færdighedstest. Denne vejledning bør være offentlig tilgængelig eller kunne udleveres også til forældre (jf. *anbefaling nr. 38*).

Argumenter for at følge

- I. Når der ikke bruges unødigt tid på at forstå, hvad testresultaterne betyder, frigøres lærerne til det egentlige pædagogiske og formative arbejde med eleverne om resultaterne.
- II. Det vil både lette lærerens arbejde og facilitere en bedre samtale med forældre, at der anvendes den samme fremstilling af elevens resultater til forældrene som læreren selv anvender pædagogisk, idet læreren gennem sin egen brug af materialet får en grundig indsigt i resultater og den måde, de er fremstillet på.

Argumenter for ikke at følge

- I. En sådan oversigt fremmer en opfattelse af, at det er vigtigt at sammenligne elever i klassen med hinanden.
- II. En sådan oversigt giver indtryk af, at det er let og hurtigt at forstå og fortolke resultaterne. Det er det ikke – der er tale om test af færdigheder, som det kræver indsigt og opmærksomhed at fortolke, og det tager tid at vurdere, hvad testen siger om hver enkelt barns faglige niveau.

ANBEFALING 40

Det anbefales at afprøve overskuelige og informative former for tilbagemelding om testresultater til lærere, hvor man i højest mulig grad udnytter de fordele som lineære test giver i forhold til lærernes direkte indsigt i de opgaver, som en hel klasse har fået.

²⁷ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 2* og *48*.

Formuleringerne knytter sig til få, centrale færdigheder. Der benyttes formuleringer af følgende type: Opgaverne 3, 5 og 7 handler om brøkgregning. Du har i din klasse en gruppe af elever (nemlig A, B, C og D), som udviste særlig dygtighed i disse opgaver. Du har også en gruppe af elever (nemlig E, F, G og J), som har særlig udfordringer med disse opgaver.

Lærerne modtager sådanne formuleringer direkte sprogligt, eventuelt suppleret med en visuel oversigt. Lærerne kan klikke på information om det underliggende beregningsgrundlag for disse udsagn, hvis de ønsker det.

ANBEFALING 41

Det anbefales, at formidlingen af testresultater til elever og forældre er læringsorienteret og dialogbaseret.

Formidlingen af testresultater til elever og forældre bør være læringsorienteret og dialogbaseret og den bør kobles til, hvad læreren/lærergruppen i øvrigt ved om elevens/klassens niveau og udfordringer.

Afvikling af test og formidling af resultaterne kan lokalt tænkes ind i eksisterende samtale- og møderutiner. Der er en risiko for, at når først testresultatet formidles, kan både elever, forældre og lærere glemme, hvad de ellers ved om elevens progression, og alene fokusere på testresultatet. Med sigte på at undgå dette anbefales det, at der udvikles en eller flere modeller for formidling og brug af resultaterne. Fx kunne der arbejdes på, at lærere tilbydes et redskab, som forløber i en tretrinsraket:

1. Inden resultatet præsenteres, angiver læreren, hvor læreren vurderer eleven er i sin faglige udvikling (her kan testens kriteriebaserede skala benyttes, hvis den findes).
2. Derefter modtager læreren resultatet af testen.
3. Hvis der er uoverensstemmelse mellem (1) og (2), må læreren overveje, om det er testen, der var misvisende eller lærerens vurdering, der var misvisende (eller en mellemting). I (sjældne) tilfælde med særlig stor diskrepans kan læreren overveje at tage en ny eller anden test med eleven.

I formidling af resultater til forældre eller i samtaler med elever om resultatet kan samme tretrinsraket benyttes.

I tilknytning til *anbefaling nr. 66* om eksperimenter og følgeforskning anbefales det, at denne og andre modeller for formidling afprøves, inden de iværksættes.

Det anbefales, at denne og lignende modeller stilles til rådighed for lærerne som et tilbud, men ikke som et krav.

Argumenter for at følge

1. Ambitionen er, at sådanne redskaber til formidling af resultater vil hjælpe elever/forældre/lærere til at indse, at de på den ene side har kvalificerede vurderinger af elevens faglige udvikling på forhånd - og på den anden side, at de nye testresultater er en anledning til at stoppe op og overveje, om eleven måske er dygtigere, end man gik og troede (og måske kan have glæde af flere udfordringer), eller måske ikke er så langt endnu (og derfor har brug for mere støtte) - eller om testresultat formentlig var misvisende, da det stred mod megen anden viden om eleven.

Argumenter for ikke at følge

1. Skoler kan lokalt være udfordret ved at have meget lidt mødekapacitet, som bruges på forældresamtaler. Det er en udfordring, hvis testresultater fra Folkeskolens Nationale Færdighedstest får hovedfokus på den begrænsede tid, der er til forældresamtaler.

- II. Tidsforbruget på tilbagemeldinger af testresultat til skole-hjem-samtalerne kan blive uhen-sigtsmæssigt højt.
- III. Denne anbefaling betyder, at der pålægges tidsforbrug som forhindrer, at tiden bruges på det, der anses for hensigtsmæssigt lokalt.

ANBEFALING 42

Det anbefales, at elevens resultater ikke formidles direkte og ubearbejdet til forældrene. Resulta-terne inddrages i skolens samlede, regelmæssige underretning af forældrene om elevens udbytte af skolegangen, jf. folkeskoleloven § 13.

Hvis forældrene får et officielt udseende papir med elevens testresultat, som i givet fald kan være det eneste forældrene ser på skrift, kan det vanskeliggøre at forklare forældrene, at testresultatet kun er ét parameter i den professionelle vurdering af elevens udbytte af undervisningen. Dermed bliver der for stor risiko for forældrepres, som kan bevirke den uhen-sigtsmæssige *teaching to the test*.

Argumenter for ikke at følge

- I. Forældrene bør modtage oplysninger om elevens testresultater skriftligt og i et format, der ikke er defineret af læreren. Det behøver ikke være alle oplysninger.
- II. Forældrene skal have mulighed for at tilgå resultatet uden læreren som fortolkningsled.
- III. Formidling af testresultater vil komme til at fylde uforholdsmæssigt meget på skole-hjem-samtaler. Ofte er der kun 15 minutter til rådighed til den enkelte elev.

ANBEFALING 43

Det anbefales, at det sikres, at data fra Folkeskolens Nationale Færdighedstest kan følge ele-verne gennem hele deres skoleforløb, uanset skift af skole eller skoleform.²⁸

Argumenter for at følge

- I. Lærere, der sent i skoleforløbet modtager elever i forbindelse med skoleskift, kan danne sig et overblik over elevens historik.

Argument mod at følge

- I. Den enkelte elev bør kunne starte i nyt skoleforløb uden, at tidligere faglige resultater følger med til læreren, da det kan give et "forkert" forhåndsindtryk, da dårlige testresultater kan have mange andre årsager end lavt niveau af færdigheder, og dette vil dokumenteres på anden vis.
- II. Det er uklart, om børns rettigheder i forhold til GDPR overholdes ved at gøre dette.

ANBEFALING 44

Det anbefales, at testsystemets tilbagemeldinger udformes på en måde, så det er muligt at over-føre den del af testresultaterne, som viser elevens optimale udviklingszone direkte til *meddelel-sesbogen*, hvis skolen vælger det.

En måde at gøre dette kunne være at gøre data tilgængelige for leverandører af it-systemer, der dan-ner rammen om skolernes *meddelelsesbøger*, på en måde så data, af lærerne, kan tilvælges som indhold i *meddelelsesbogen*.

²⁸ Tine Nielsen støtter ikke *anbefaling nr. 43*.

Argumenter for at følge

- I. På den måde tillægges elevens udviklingsmuligheder mest vægt.
- II. Det vil lette overblikket over elevens testresultater gennem hele skolegangen for både elever, forældre og lærere.
- III. Det vil gøre det enkelt for elever og forældre at finde testresultatet.
- IV. På den måde undgås behovet for at skabe en ny portal, som har til formål at formidle testresultaterne.
- V. Det sikres, at elever og forældre nemt har adgang til resultater fra Folkeskolens Nationale Færdighedstest, og giver samtidig mulighed for at Folkeskolens Nationale Færdighedstest indgår som en del af arbejdet med *meddelelsesbogen*.

Argumenter for ikke at følge

- I. Resultaterne fra nationale test, herunder fremtidige Folkeskolens Nationale Færdighedstest, er ikke lette at fortolke og forstå. De bør derfor ikke deles med forældre uden at en lærer, der har den fornødne forståelse og indsigt, hjælper med at forstå og fortolke resultaterne.

4.3. Psykometriske egenskaber

ANBEFALING 45

Det anbefales, at måleusikkerheden i de nye lineære Folkeskolens Nationale Færdighedstest sættes til *standard error of measurement (SEM)* på omkring 0,3.

Med måleusikkerhed menes *standard error of measurement (SEM)* af de person parameter estimer, som kan uddrages fra den psykometriske målingsmodel. Denne anbefales sat til omkring 0,3 (svarende til det oftest brugte stop-kriterie i adaptive test) for alle, der testes, dvs. henover hele skalaen, undtagen yderpunkterne, hvor dette ikke er muligt at opnå. Dette kræver, at sværhedsgraderne er fordelt henover hele skalaen. SEM på 0,3 på person parameter estimerne (fx Rasch-scores på logit-skala) er det niveau, der bør tilstræbes for individuel personvurdering i pædagogiske test som Folkeskolens Nationale Færdighedstest, jf. internationale standarder. Lav måleusikkerhed betragtes som essentielt, da resultater med for høj usikkerhed ikke vil have samme udsigelseskraft, og dermed heller ikke samme anvendelighed i lærerens formative arbejde, som mere sikre resultater vil have. Indenfor *item response teori*, som er det test-teoretiske paradigme, som ligger under Folkeskolens Nationale Færdighedstest, er beregning af SEM mere kompliceret end i klassisk testteori, men ikke som sådan kompliceret i forbindelse med de item-analyser, der gennemføres i Styrelsen for It og Læring (jf. fx Svend Kreiners notat om opgavetyper og usikkerhed i pædagogiske test, som findes på Børne- og Undervisningsministeriets hjemmeside²⁹).

Argumenter for at følge

- I. Målesikkerhed på elevniveau er alpha og omega, når resultaterne skal anvendes på elevniveau.
- II. Lærerne vil gerne have resultater på elevniveau, som de kan stole på, således at de kan anvende test formativt i deres arbejde med eleverne.
- III. Målesikkerheden bør ikke være dårligere end standarden for adaptive test til vurderinger på individniveau, blot fordi der er tale om lineære test.
- IV. Hvis måleusikkerheden er stor vil lærere opleve, at resultaterne for en del elever ikke passer - noget som over tid vil reducere tilliden til Folkeskolens Nationale Færdighedstest.
- V. Hvis måleusikkerheden er høj vil en beskrivelse af, hvad eleven kan eller er tæt på at kunne, være fejlagtig for en stor andel elever.

²⁹ Kreiner, S. (2017). *Om opgavetyper og usikkerhed i de nationale test*: <https://www.uvm.dk/-/media/filer/uvm/udd/folke/pdf17/sep/170913-om-opgavetyper-og-usikkerhed-i-de-nationale-test.pdf> (jf. også anbefaling nr. 7 og anbefaling nr. 46).

Argumenter for ikke at følge

- I. Udvidede rutiner for afprøvning af opgaver kommer oplagt med en ekstra udgift.
- II. Der er en tendens til, at man fokuserer på én egenskab ved test ad gangen. Det kan fx være præcisionen, tidsforbruget, antallet af målte færdigheder eller testprincippet. Det er uhensigtsmæssigt. I stedet bør man foretage en samlet afvejning af de forskellige hensyn i et helhedsperspektiv.

ANBEFALING 46

Det anbefales, at Folkeskolens Nationale Færdighedstest har god indre konsistens/reliabilitet på mindst 0,85.

Validering vil i korte træk sige, at man samler og vurderer evidens for at støtte helt specifikke tolkninger af testscores. For Folkeskolens Nationale Færdighedstest er den helt grundlæggende tolkning, at elevens scorer på testene repræsenterer elevens dygtighed inden det faglige område, som testens rammeværk definerer. Man vil derfor ønske at udvikle test, der på den ene side ikke måler for lidt af det, som skal måles, og på den anden side ikke måler for meget af noget, som ikke skal være en del af testen - som i engelsk faglitteratur gerne refereres til som henholdsvis *construct underrepresentation* og *construct irrelevant variance*. Et krav om reliabilitet på mindst 0,85 er erfaringsmæssigt et kriterium for testscorer, som repræsenterer en rimelig afvejning af disse hensyn, og som samtidig tager hensyn til *anbefaling nr. 45*, som omhandler målesikkerhed på elevniveau.

Argumenter for at følge

- I. Sammen med flere af de andre anbefalinger vil et krav om reliabilitet være central og nødvendig evidens for at støtte et argument om, at Folkeskolens Nationale Færdighedstest har tilfredsstillende begrebsvaliditet.

Argumenter for ikke at følge

- I. Der er en tendens til, at man fokuserer på én egenskab ved test ad gangen. Det kan fx være reliabilitet, tidsforbruget, antallet af målte færdigheder eller testprincippet. Det er uhensigtsmæssigt. I stedet bør man foretage en samlet afvejning af de forskellige hensyn i et helhedsperspektiv.

ANBEFALING 47

Det anbefales, at testsystemet til Folkeskolens Nationale Færdighedstest er *open source* og lever op til åbne standarder for opgavedesign.

Den internationale non-profit-organisation 1EdTech (tidligere IMS Global) står bag åbne standarder for udvikling af opgaver til brug i test. Det betyder, at man ikke bindes til ét system, men kan skifte. Det betyder også, at man kan udveksle opgaver med andre aktører (andre lande, nationale og internationale organisationer mv.).

Det sikrer samtidig, at testopgaverne kan være *state of the art*, både teknisk og designmæssigt, fx muliggøre integration af interaktive apps (geometriprogrammet GeoGebra, regnearksprogrammer, programmer til visuel programmering, programmer til (også elevens) design af hjemmesider, præsentationer, video, animationer osv.).

Argumenter for at følge

- I. Der findes et fungerende *open source system*, TAO (Testing Assisté par Ordinateur), som anvendes af en lang række undervisningsministerier (Frankrig, Portugal, Italien, Luxembourg, Norge, Brasilien og mange flere).

- II. Der er fx i regi af den internationale forening FLIP+ initiativer til at skabe fælles banker af opgaver (itemlibrary.org).
- III. Hvis man knytter sig til en privat aktør med *closed source* og som ikke benytter åbne standarder, er man låst til denne udbyder.

Argumenter for ikke at følge

- I. Opgaveformater og opgavetyper, som eleverne møder i skolen, kan være begrænsende for fagmiljøerne, som skal lave Folkeskolens Nationale Færdighedstest. Et implicit formål for Folkeskolens Nationale Færdighedstest bør også være, at de skal fungere som forbilleder for lærernes egne vurderinger, og på den måde kan det være relevant, at testene repræsenterer innovation og nytænkning om, hvordan opgaver kan udformes.

ANBEFALING 48

Det anbefales, at testsystemet til Folkeskolens Nationale Færdighedstest designes og udvikles som en testbank (altså både en opgavebank, men også en bank af hele test/opgavesæt), der kan dække testgennemførelsen i flere år, der kan linkes mellem klassetrin, der er mulighed for frivillig gentestning senere på skoleåret, og hvor allerede anvendte opgaver kan anvendes til gentestning.³⁰

Pædagogisk arbejde med den enkelte elev på basis af testresultater kræver, at læreren har kendskab til de færdighedsområder, testen dækker, før testen gennemføres, og at læreren efter testen er gennemført, har kendskab til testens konkrete indhold, herunder i arbejdet med resultaterne. Et andet væsentligt parameter for lærernes anvendelse af testresultater til pædagogiske formative formål er muligheden for at kunne gennemføre en test igen frivilligt senere på året.

Til monitoreringsformål vil der sædvanligvis være behov for, at opgaver ikke er offentligt kendte. Hvis opgavernes konkrete indhold først frigives, når testperioden for gennemførelsen af testen er udløbet, vil testene dog stadig kunne anvendes til monitorering. Hvis tilbagemeldinger desuden først er tilgængelige ved en testperiodes udløb, vil opgaveparametre kunne genberegnes for at sikre, at disse ikke afviger fra udviklingsfasen.

På tværs af lærernes pædagogiske anvendelse og monitoreringsformål, er det desuden væsentligt at kunne følge udviklingen i færdigheder over tid. Lærere kunne drage nytte af at kunne følge en klasses udvikling henover en række klassetrin, mens offentlig monitorering kunne drage nytte af at kunne følge en skoles, en kommunes eller alle elevers udvikling henover klassetrin. Der findes løsninger, der kan tilgodese begge disse formål. Et forslag er beskrevet kort og illustreret nedenfor:

- Til start udvikles en testbank, hvor der er test/opgavesæt til alle de klassetrin, der skal testes, således at disse kan linkes henover klassetrin med det samme. Udviklingen af testbanken indebærer samtidig, at der udvikles test/opgavesæt til hvert klassetrin til de følgende to eller tre skoleår, sådan at der indledningsvist i testbanken er udviklet obligatoriske test/opgavesæt til alle klassetrin til i hvert fald tre skoleår. Derudover udvikles til alle klassetrin også frivillige gentest, der er parallelle til de obligatoriske test, således at scores kan sammenlignes (jf. figur).
- Når Folkeskolens Nationale Færdighedstest påbegyndes i skoleåret 2026/27, kan udviklingen af test til næste runde af to til tre skoleår påbegyndes. For eksemplets skyld antager vi, at der til en start er udviklet test til tre skoleår (dvs. skoleårene 2026/27, 2027/28 og 2028/29). Ved udviklingen af test til de efterfølgende tre skoleår (dvs. skoleårene 2029/30, 2030/31 og 2031/32) kan der spares ressourcer, da der på hvert klassetrin, allerede vil være seks test/opgavesæt, hvor opgaverne vil være tilgængelige, når anden runde går i gang i skoleåret 2029/30.

³⁰ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 2* og *39*.

Disse test/opgavesæt kan således anvendes til at lave frivillige gentest til anden runde, fx ved at fordele opgaverne anderledes på de enkelte test, således at det sikres, at de er parallelle til de nye obligatoriske test/opgavesæt.

- Denne designproces kan fortsættes over tid. Efter den første udvikling skal der således kun udvikles nye opgaver til de obligatoriske test, mens alle de frivillige gentest kan genereres ud fra "gamle" opgaver i en akkumulerende opgavebank. Frivillige gentest til formative formål kan således gennemføres uden store udgifter.

Illustration af et løsningsforslag til udvikling af testbank til Folkeskolens Nationale Færdighedstest



Beskrivelse af illustrationen: I de færdighedsområder, hvor der er obligatoriske test, udvikles to sideordnede lineære test til et klassetrin (henholdsvis en obligatorisk test og en frivillig gentest), og scores på disse equates. Derudover indgår der et lille antal opgaver i de obligatoriske test (), fx en til tre opgaver, som går igen i hvert "klassetrinspar" (i illustrationen er 2. og 4. klasse et par og 4. og 6. klassetrin et par), således at disse kan linkes sammen. Optimalt set holdes disse få linking-opgaver hemmelige, da de tjener en særlig funktion. Det anses ikke for nødvendigt at have linking-opgaver i de frivillige gentest.*

Fremgangsmåden kan også anvendes ved udvikling af de frivillige Folkeskolens Nationale Færdighedstest – her udvikles således blot to frivillige test på hvert klassetrin.

Ovenstående er ét forslag til en løsning, og der kan findes andre løsninger, end det her foreslåede.

Argumenter for at følge

- I. Tilgodeser læreres behov for at kende de konkrete opgaver for at kunne arbejde formativt (når resultaterne foreligger).
- II. Tilgodeser muligheden for at kunne genteste en klasse frivilligt senere på skoleåret, med en ny og sammenlignelig test.
- III. Tilgodeser, at data kan bruges til monitoreringsformål, da opgaverne ikke er kendte på testgennemførelsestidspunktet.
- IV. Tilgodeser behov for at følge udvikling på klasseniveau (lærere), på skoleniveau (skolerne), på kommuneniveau og på nationalt niveau, henover klassetrin.

Argumenter for ikke at følge

- I. At lave en sådan testbank vil kræve store ressourcer i opstartsfasen.
- II. Det er en risiko for, at opgaveparametre fra udviklings- og opgavefasen ikke er stabile. Derfor bør opgaveparametrene genberegnes ved afslutning af testperioden, inden tilbagemeldinger genereres.

ANBEFALING 49

Det anbefales, at tilbagemeldinger om elevernes resultat i Folkeskolens Nationale Færdighedstest sker i relation til færdighedsudviklingsbeskrivelser (proficiency scales).

En test med fagligt varierede opgaver kan bruges til at beskrive den typiske faglige udvikling, og elevernes resultater kan beskrives fagligt på baggrund af en sådan *kompetence- eller færdighedsudviklingsbeskrivelse*. Tilbagemeldingen bliver derved relateret til det faglige område, ikke direkte til andre elevers resultater.

Tilbagemeldingen til lærerne kan ske i forhold hertil og kan give lærerne besked om, hvad elever med det niveau typisk er i færd med at tilegne sig af færdigheder, hvad de typisk allerede kan, og hvilke faglige udfordringer, der stadig er uden for rækkevidde. Der kan på den baggrund også gives forslag til måder at give differentieret *feed-forward* og *feedback*.

Argumenter for at følge

- I. Faglige tilbagemeldinger er en forudsætning for lærerens pædagogiske brug af test.
- II. Tilbagemelding til forældre om deres barns aktuelle faglige udvikling kan foregå på et meningsfuldt og nuanceret grundlag.
- III. Forskere og embedsmænd, der bruger testene, kan anvende de faglige udviklingsbeskrivelser til at forstå og formidle, hvad elever i en given gruppe typisk vil være i færd med at lære.
- IV. Læremiddeludviklere kan anvende de empirisk funderede færdighedsudviklingsbeskrivelser til at udforme fagligt målrettede læremidler.
- V. Færdighedsbeskrivelser giver grundlag for faglig udvikling blandt lærere, da de kan bidrage til at skabe et fagsprog om det pågældende faglige område.
- VI. Mens en normrelateret score giver eleven tilbagemelding om, hvorvidt eleven er bedre eller dårligere end andre elever, og derved opfordrer til sammenligning og oplevelse af nederlag, angiver en færdighedsudviklingsbeskrivelse, hvad eleven aktuelt er i færd med at tilegne sig af færdigheder, og hvad det giver mening at arbejde med i den kommende tid.
- VII. Beskrivelserne vil ikke være omfattende og vil derfor ikke fratage læreren eller andre det professionelle ansvar for at vurdere elevens samlede faglige niveau.
- VIII. Nogle elever vil svare på opgaverne på en måde, der statistisk er meget usandsynlig (fx rigtigt på svære opgaver, men forkert på lette). Sådanne elever kan man vælge ikke at formulere færdighedsbeskrivelser for.
- IX. Der er en meget lang forskningstradition for udvikling af proficiency scores eller kriteriebase-rede resultater – helt tilbage fra Glaser (1963³¹). Der er således solidt forskningsmæssigt belæg og tradition for at udarbejde og anvende færdighedsbeskrivelser.

Argumenter for ikke at følge

- I. Tilbagemeldinger i form af færdighedsbeskrivelser kan begrænse forståelsen af eleven hos læreren, forældrene og eleven selv. Jo mere detaljerede beskrivelserne bliver, jo mindre overlades til lærerens vurdering. Det sætter store krav til, at færdighedsbeskrivelsen (og angivelsen mulige udviklingsstier) er korrekt og relevant. Det gør det sværere for læreren at indføre sin egen vurdering af elevens færdigheder og udviklingsmuligheder.
- II. Der findes kun i begrænset omfang forlæg for individuelle tilbagemeldinger i form af færdighedsbeskrivelser, fx er PISA/PIRLS ikke designet til at udtale sig om enkeltelever. Det er derfor svært at vurdere, i hvilket omfang det vil være muligt at udvikle gyldige tilbagemeldinger inden for den bredde vifte af færdighedsområder og klassetrin som Folkeskolens Nationale Færdighedstest skal dække.

³¹ Glaser, R. (1963). *Instructional technology and the measurement of learning outcomes*. *American Psychologist*. 18 (8), s. 519–522.

- III. Færdighedsbeskrivelser har ligesom skalascorer en underliggende skala, hvis pålidelighed og gyldighed kan vurderes. Men de enkelte udsagn i færdighedsbeskrivelser kan også være mere eller mindre pålidelige/gyldige, så den psykometriske vurdering af færdighedsbeskrivelser bør være mere omfattende.
- IV. Det kræver raffineret kommunikation over for enkelteleven, hvis færdighedsbeskrivelsen er udviklet til at gælde den typiske elev på et bestemt færdighedstrin.
- V. Færdighedsbeskrivelser kan stadig være skuffende for eleven, og man kan stadig risikere sammenligninger mellem elever.

ANBEFALING 50

Det anbefales, at der gennem kontinuerlig afprøvning af opgaver i udviklingsfasen sikres, at der er god tilpasning mellem opgavernes sværhedsgrader og elevernes færdighedsniveauer, således at alle færdighedsniveauer dækkes og alle elever gives samme muligheder for at opnå et fyldestgørende testresultat (jf. også anbefaling nr. 9 og anbefaling nr. 10).³²

Ses der tilbage på de tidligere adaptive nationale test i folkeskolen, blev det i den seneste evaluering dokumenteret³³, at der var flere testområder, hvor der manglede opgaver, som passede til de dygtigste elever. Det betød, at disse elevers niveau reelt ikke blev afprøvet, og at deres resultater ikke var valide. Med nyudviklede opgaver kan dette i princippet også ske i den lave ende af færdighedsskalaerne. Det er derfor nødvendigt at sikre, at der er opgaver, der modsvarer alle færdighedsniveauer i Folkeskolens Nationale Færdighedstest. På den måde er det muligt at opnå god *targeting* og tilstrækkelig lav måleusikkerhed til, at alle elever kan blive testet på deres niveau, og opnå et tilstrækkeligt præcist og fyldestgørende resultat (jf. også *anbefaling nr. 10* og *anbefaling nr. 45*). Det er desuden væsentligt for elevernes oplevelse af testsituationen, at der er "opgaver til alle". Hvis der mangler opgaver i den lave ende, vil de svageste elever få en unødigt negativ oplevelse, mens en mangel på opgaver i den høje ende vil betyde, at de dygtigste elever, ud over at kede sig, heller ikke får en reel afprøvning af deres færdigheder (jf. også *anbefaling nr. 14*, *anbefaling nr. 16* og *anbefaling nr. 17*).

I praksis kan dette løses ved at sammensætte test, der indeholder opgaver på så mange sværhedsgradsniveauer som muligt, og sørge for, at også de laveste og højeste niveauer er tilstrækkeligt dækket.

Argumenter for at følge

- I. Alle elever gives mulighed for at opnå et præcist testresultat, som er retvisende i forhold til deres færdighedsniveau.
- II. Lærernes muligheder for at arbejde formativt med eleverne baseret på deres resultater forbedres, idet der er præcise og retvisende resultater til elever på alle niveauer.
- III. Alle elever har ret til lige vilkår i forhold til at kunne gennemføre en test, uanset om de er blandt de svageste, i midterfeltet, eller de dygtigste inden for et færdighedsområde. Denne anbefaling gør det muligt at undgå forskelsbehandling af de svageste og de dygtigste elever.

³² Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr. 7, 8, 9, 10, 45, 46, 51* og *positionsrapport fra Rolf Vegar Olsen med flere*.

³³ Styrelsen for It og Læring (STIL). (2020). *Evaluering af de statistiske aspekter ved de nationale test*. STIL. Børne- og Undervisningsministeriet.

³³ Det Nationale Forsknings- og Analysecenter for Velfærd, VIVE. (2020). *Review af evalueringen af de statistiske aspekter ved de nationale test – Delrapport 1: Evaluering af de nationale test*. VIVE.

³³ Nielsen, T. (2019). *Review af evalueringen af de statistiske aspekter ved de nationale test – Delrapport 1: Evaluering af de nationale test*: [Microsoft Word - REVIEW AF DOKUMENTATIONEN FOR DE NATIONALE TEST TINE NIELSEN3.docx](#)

Argumenter for ikke at følge

- I. Mange meget lette og svære opgaver kan opleves som spild af tid eller frustrerende af det store flertal af elever, og kan have begrænset informationsværdi. Det er sandsynligvis ikke nødvendigt at kunne skelne meget fint i bunden og toppen af skalaerne, og derfor kan det være en fordel at have et flertal af opgaver omkring den brede midte.
- II. Hvis opgavesættet indeholder ekstra mange opgaver, der er meget lette og meget svære i forhold til klassetrinnets gennemsnitlige niveau, vil den store majoritet af eleverne skulle besvare mange opgaver, som er for lette (og måske kedelige), og mange som er for svære (og måske nedslående) for dem. Det er for stor en ulempe i forhold til hensynet til at kunne skelne mellem elever, som er fagligt meget dygtige/udfordrede og elever, som er fagligt endnu mere dygtige/udfordrede. Hvis den enkelte lærer har brug for at kunne foretage denne skelnen, kan læreren teste de pågældende elever med en supplerende test. Hvis dette fx er muligt via en opgavebank, som opgavesæt kan trækkes fra jf. *anbefaling nr. 48*, vil det være let at tilbyde supplerende opgavesæt, som er relativt svære eller lette i forhold til klassetrinnets almindelige opgavesæt.
- III. Det har ikke store praktiske konsekvenser for det pædagogiske arbejde, hvis Folkeskolens Nationale Færdighedstest ikke er i stand til at skelne finkornet mellem meget dygtige og særdeles dygtige elever. Til sammenligning rummer fx karakterskalaen, benyttet ved folkeskolens prøver, kun to karaktertrin (10 og 12) til de allerdygtigste. Derfor er det svært at forstå, at der skal stilles mere vidtgående krav til Folkeskolens Nationale Færdighedstest end til folkeskolens prøver.
- IV. Modargument I til III vil resultere i, at elever i toppen og bunden får resultater, som kan være meget langt fra sandheden. Særligt for de svageste elever kan det være et alvorligt problem, da de ikke får de indsatser eller den opmærksomhed, de har behov for. De stærkeste kan få unødvendige oplevelser af nederlag, når de sammenligner sig med andre elever, som har klarlagt sig bedre, men reelt er dårligere.

ANBEFALING 51

Det anbefales, at testresultaterne fra Folkeskolens Nationale Færdighedstest rapporteres på en måde, som støtter kriteriebaseret scoring og fortolkning. Fortolkningerne/beskrivelserne af testresultaterne bør bestå af positive udsagn (hvad eleven kan/mestrer/forstår), samt udsagn om hvad, der kunne være elevens proximale udviklingszone (hvad eleven er nær ved at kunne/mestre/forstå).³⁴

Kriteriebaseret scoring og fortolkning har den fordel, at der kan gives et rimeligt præcist billede af, hvad den enkelte elev har svært henholdsvis ikke har svært ved inden for et givent færdighedsområde eller delområde. Kriteriebaseret scoring kan således danne grundlag for en formativ pædagogisk anvendelse af testresultaterne for den enkelte elev. Kriteriebaserede skalaer består af fagligt funderede kriteriebeskrivelser for forskellige resultatniveauer, og den enkelte elevs resultat vurderes i forhold til disse kriterier. Kriterierne giver objektive standarder, som ikke ændres over tid.

De kriteriebaserede tilbagemeldinger kan genereres automatisk af testsystemet, så snart en elev har gennemført testen, og det er ikke nødvendigt at afvente elevpopulationens resultater for at få årets norm og resultaterne beregnet (jf. *anbefaling nr. 7*).

³⁴ Forfatteren ønsker, at der i relation til denne anbefaling henvises til *anbefaling nr., 9, 10, 17, 50* og *positionsrapport fra Rolf Vegar Olsen med flere*.

Med kriteriebaseret scoring er det også muligt at "mappe" en hel klasses resultater i et grafisk overblik til læreren, medfølgende en standardiseret beskrivelse af, hvad de typiske vanskeligheder er på de forskellige scoreniveauer. På den måde kan læreren have et overblik over klassens resultater til egen brug i forbindelse med tilrettelæggelse af den videre undervisning.

Kriteriebaseret scoring til pædagogisk brug og formidling af resultater til elever, lærer og forældre, kan suppleres med normbaseret scoring til monitoreringsformål, fx på kommuneniveau eller nationalt niveau.

Argumenter for at følge

- I. Elevernes trivsel vil sandsynligvis øges, når det sociale sammenligningspres mindskes og der fokuseres på den enkelte elevs resultater i sin egen ret.
- II. Læreren pædagogiske arbejde med testresultaterne som formativ evaluering af færdigheder styrkes.
- III. Forældres muligheder for konkret støtte i forbindelse med lektielæsning øges, når testtilbage-meldingen beskriver, hvilke områder deres barn i dets egen ret har vanskeligheder i, og hvilke der er uproblematisk.
- IV. Kriteriebaserede scores har desuden den fordel, at de indeholder information om, hvad eleven har svært ved/ikke har svært ved som en del af opgørelsen. Det er således ikke nødvendigt for andre end læreren at kende det konkrete opgaveindhold eller de forskellige opgaveområder. Hvis der skrives en tilstrækkelig god kriterieforklaring, vil det også være muligt at arbejde formativt med elever, som har gennemført hemmeligholdte opgaver, så længe opgaverne først tilgår læreren, når testperioden er afsluttet.
- V. Der vil være tale om fagligt baserede og vedtagne kriterier for, hvad det vil sige at have svært ved eksempelvis et område i matematik/ikke have svært ved et område i matematikken, som anvendes for alle elever og derfor giver en beskrivelse af, hvor eleven ligger i forhold til kriterierne (*jf. argument I mod nr. III for modsætningen*).
- VI. Det er fuldt muligt at beskrive nationale og andre grupperingers fordelinger ved hjælp af kriteriebaserede scores, og det er muligt at standardisere sådanne scores og beregne normbaserede fordelinger i forbindelse med forskning eller national monitoring, hvis det ønskes. Det er således unødvendigt at beleme elever, lærere og forældre med normbaserede resultater.

Argumenter for ikke at følge

- I. At udvikle præcise og gode kriteriebeskrivelser er arbejds- og ressourcekrævende.
- II. Det er nemmere at udvikle præcise og didaktisk meningsfulde beskrivelser for test, som måler et relativt smalt begreb/område. For test af relativt bredere begreb/områder bliver beskrivelserne af elevernes kompetence nødvendigvis også mere generelle/abstrakte.
- III. I modsætning vil normbaseret scoring blot fortælle, hvor eleven ligger i forhold til de andre elever (normen). Hvis normen ligger lavt kan en elev, der ligger lavt på den testede færdighed alligevel få en score, der viser, at eleven ligger højt, da der sammenholdes med normen – det omvendte gør sig også gældende. Dermed får læreren ikke information om, hvor eleven reelt set ligger.
- IV. Der findes kun i begrænset omfang forlæg for individuelle tilbagemeldinger i form af kriterie-opfyldelse. Det er derfor svært at vurdere, i hvilket omfang det vil være muligt at udvikle gyldige tilbagemeldinger inden for den brede vifte af kriterier og klasstrin som Folkeskolens Nationale Færdighedstest skal dække.
- V. Selvom færdighedsbeskrivelser kan omgå en skalascoretilbagemelding og de umiddelbare problemer med sammenligning af scorer, så kan færdighedsbeskrivelsen stadig være skuffende for eleven, og man kan stadig risikere sammenligninger mellem elever.
- VI. Inden for mere komplekse aspekter af læsning (læseforståelse og fortolkning) er det vanskeligt at angive motiverede kriterier, der har gyldighed ud over selve testen. Fx afhænger muligheden for at kunne reflektere over en tekst, drage følgeslutninger eller sammenholde informationer på tværs af tekster helt af teksternes indholdsmæssige sværhedsgrad. Det er derfor svært at angive, hvad man kan forvente på et givet klasstrin, og hvordan det adskiller sig fra

andre klassetrin. Man kan eventuelt forankre kriterier i tekster passende for klassetrinnet. Det vil dog være udfordret af, at der ikke findes klare standarder for, hvad der er passende til klassetrinnet, og der er stor variation i praksis.

- VII. Et udsagn om en elevs proximale udviklingszone kan gælde en typisk elev (begrundet ud fra teori eller empiriske tendenser), men behøver ikke at gælde for det enkelte barn. Det kan føre til misforståelser, hvis testene giver indtryk af at kende det enkelte barn bedre end læreren.
- VIII. Folkeskolens Nationale Færdighedstest bør afgrænse sig til at teste få centrale færdigheder, og må gerne støtte det pædagogiske arbejde, men bør ikke give indtryk af at foreskrive en bestemt form for pædagogisk indsats.

ANBEFALING 52

Det anbefales, at elevernes scorer i tilbagemeldingen oversættes til et begrænset antal brede trin med pædagogisk orienterede betegnelser.

Scorer – uanset om de er norm- eller kriteriebaserede – er fundamentet for elevkarakteristikken, men kan opleves som uhåndgribelige og unødigt detaljerede. Det er derfor hensigtsmæssigt at oversætte til en skala med et begrænset antal intervaller, der i modsætning til tidligere bør gives pædagogisk orienterede titler i stedet for titler, der er norm- eller kriterieorienterede. Fx kunne trinnene i en tretrinsskala gives titlerne "har brug for ekstra faglig opmærksomhed" (i stedet for normbaseret "under middel" eller kriteriebaseret "mangelfuld"), "er på niveau med klassetrinnet aktiviteter" (i stedet for "middel" eller "god") og "har brug for flere eller andre udfordringer" (i stedet for "over middel" eller "rigtigt god").

Derudover kan læreren eventuelt tilbydes en grafisk fremstilling af elevens scorer i forhold til trinnenes grænser for at give indtryk af elevens placering på et trin. Sådanne fremstillinger bør holdes enkle og afprøves i målgruppen.

Argumenter for at følge

- I. Lærere vil nok have begrænset mulighed for at kunne håndtere meget mere end tre niveauer i klassen.
- II. En trinplacering uden detaljeret færdighedsbeskrivelse overlader den nærmere karakteristik og opfølgning til skolens fagpersonale, dvs. tæt på eleven.
- III. En trinplacering kan eventuelt suppleres med en mere uddybende færdighedsbeskrivelse.
- IV. De præcise grænser for trininddeling er delvist vilkårlige. Derfor er det nyttigt, at elevens afstand til nærmeste grænse visualiseres.
- V. Scorer er velkendte fra andre test og veletablerede i psykometri og testudvikling, og de er relativt enkle at beregne og pålidelighedsvurdere.
- VI. Scorer kan sammenholdes med resultater fra andre test, fx hvis læreren er usikker på resultatets gyldighed.

Argumenter for ikke at følge

- I. Scorer og trinplaceringer gør det nærliggende at sammenligne testresultater elever imellem. Det kan føre til uønsket konkurrence og fokus på en trinplacering i stedet for faglig udvikling.
- II. Scorer og trinplaceringer fortæller ikke noget kvalitativt om elevens færdighed, og giver ikke direkte anvisninger til opfølgning eller udvikling.
- III. Det kan være udfordrende at kommunikere til en elev, at denne ligger på et lavere trin end eleven forventede. Uheldig kommunikation kan give nederlagsoplevelse.
- IV. En score giver forældre incitament til at kræve, at læreren bringer deres børn "op på niveau", uden at der er grundlag for at tale om, hvad det vil sige.
- V. En kategorisering af scores i intervaller vil have forskellige betydninger alt efter, om de baseres på normbaseret scoring eller kriteriebaseret scoring. I en normbaseret udgave vil "har brug for ekstra faglig opmærksomhed" blot forholde sig til, hvor eleven ligger i forhold til (den nationale) norm. Således vil dette kunne betyde, at eleven er god til matematik, hvis normen ligger

lavt (matematik er blot brugt som eksempel). Det vil kunne betyde, at eleven er dårlig til matematik, hvis normen ligger højt. Det er således uden værdi for læreren, da det reelle matematikniveau ikke kendes, og normen vil flytte sig over tid. I en kriteriebaseret udgave vil udsagnet "har brug for ekstra faglig opmærksomhed" være båret af fagligt definerede kriterier for, hvad der skal til for at være god henholdsvis dårlig til matematik, og læreren vil vide, hvad elevens matematikniveau er. Tillige kan eleverne sammenlignes på deres reelle niveau (jf. også *anbefaling nr. 51*).

ANBEFALING 53

Det anbefales, at eksisterende opgavemateriale fra tidligere og nuværende nationale test vurderes for genbrug i Folkeskolens Nationale Færdighedstest i den grad de tilfredsstillende nye krav, og kan implementeres i det nye testsystem.

Det findes allerede et righoldigt bibliotek af opgaver i de eksisterende nationale test. Man bør på et tidligt stadie lave en systematisk gennemgang af dette opgavemateriale med henblik på at anvende eller videreudvikle disse opgaver til brug i Folkeskolens Nationale Færdighedstest. Det er imidlertid vigtigt at præcisere, at alle opgaver til de nye test skal piloteres på ny, for at teste om de tilfredsstillende alle krav, jf. andre anbefalinger, fx krav knyttet til muligheden for didaktisk anvendelse. Der vil endvidere være behov for at tilpasse opgaverne til den nye IT-løsning.

Argumenter for at følge

- I. At udvikle gode opgaver til faglige test er tidkrævende og dyrt. Der er al mulig grund til at antage, at den eksisterende opgavebank har indhold, som i en eller anden form kan benyttes fremadrettet.

Argumenter for ikke at følge

- I. Det er en bekymring, baseret på indsigt i det gældende opgavemateriale, at mange af opgaverne i fra de tidligere og nuværende nationale tests opgavebank ikke egner sig godt til de nye Folkeskolens Nationale Færdighedstest.
- II. Folkeskolens Nationale Færdighedstest skal måle andre færdigheder og med andre opgaveformater, og det vil derfor tage alt for lang tid først at udvælge og dernæst at tilpasse og implementere opgaver fra de tidligere og nuværende test.

ANBEFALING 54

Det anbefales, at der i opbygningen af opgaverne til Folkeskolens Nationale Færdighedstest arbejdes med passende repræsentationsformer, således at tekst og illustrationer er relevante for den stillede opgave.

Eleverne skal derfor kun præsenteres for relevante billeder/diagrammer/animationer/simulationer i opgaverne således, at de ikke misforstår opgaverne. Fx er det med et simpelt eksempel uhensigtsmæssigt at have billeder af en båd i en opgave, der handler om biler, eller at have irrelevante animationer m.v. Der bør altså undgås såkaldt *dressing up* af opgaver medmindre, at det er nødvendigt af hensyn til at teste elevens evne til at identificere relevant information blandt en mængde irrelevant information.

Argumenter for at følge

- I. At opgavernes problemstillinger fremstår tydeligt for eleverne.
- II. At eleverne kun skal forholde sig til opgaven og ikke til ikke relevante informationer.
- III. Det mindsker muligheden for, at opgaverne bliver misforstået af eleverne, og at eleverne forvirres.

ANBEFALING 55

Det anbefales, at der både er opgaver, som inddrager hverdagsprog og opgaver, der inddrager fagsprog, så forskellige færdighedsniveauer tilgodeses.

Afhængig af elevens baggrund kan fagsproget være mere eller mindre udviklet. En elevs besvarelse af testen giver derfor ikke nødvendigvis indblik i elevens færdigheder i at reflektere over fx undersøgelser og modeller eller give bud på sammenhænge ved hjælp af fagsprog. Eleven kan måske besvare opgaven med hverdagsprog.

Argumenter for at følge

- I. Testen skal også tage hensyn til de elever, der ikke har et fagsprog, men har en faglig viden.

ANBEFALING 56

Det anbefales at være opmærksom på, at udvikling og brug af Folkeskolens Nationale Færdighedstest nødvendigvis indebærer nogle afvejninger mellem forskellige hensyn.

Eksempler på sådanne afvejninger kan være:

- Afbalanceringen af de hensyn, der tages til de mange forskellige brugere af testresultaterne.
- Præcisionen i testresultater versus elevens tidsforbrug til at deltage i test.
- Hensynet til præcision i testresultaterne versus hensynet til praktisk anvendelse (Resultater kan være praktisk nyttige, også selv om de ikke er perfekte med hensyn til præcision).
- De ressourcer, der bruges til opkvalificering af lærere, skoleledere og andre i forhold til testforståelse versus de ressourcer, der bruges til andre formål i skolen.
- Hensynet til en redelig beskrivelse af testresultaternes tekniske og statistiske forudsætninger versus hensynet til almen forståelighed af testresultaterne blandt de mest talrige brugere (lærere, elever og forældre).

Hvis man erkender sådanne afvejninger, følger heraf, at selv den bedste praktiske udmøntning af Folkeskolens Nationale Færdighedstest vil medføre, at ikke alle hensyn tilgodeses optimalt.

Argumenter for ikke at følge

- I. Det er indlysende, at hensyn må afvejes. Men anbefalingen giver en række eksempler, som samlet antyder, at man kan slække på grundlæggende psykometriske og faglige krav, uden at det resulterer i alvorlige problemer. Det kan man ikke. En test som anvendes til at evaluere professionelles og institutioners kvalitet, skal i sagens natur selv være af høj kvalitet. Og en test som bruges til at afgøre indsatsen over for elever, skal ligeledes være af høj kvalitet.
- II. En test, som ikke tilfredsstillende helt grundlæggende krav til psykometrisk kvalitet, giver begrænset og usikker information om enkeltelever. Principper, som identificerer standarder for henholdsvis målefejl og reliabilitet, skal derfor prioriteres højere end andre hensyn (jf. *anbefaling nr. 45 og positionsrapport fra Rolf Vegar Olsen med flere*).

ANBEFALING 57

Det anbefales, at der i forbindelse med udviklingen af det nye testsystem til Folkeskolens Nationale Færdighedstest tilvejebringes data på forholdet mellem testenes målesikkerhed, tidsmæssige længde og faglige bredde, så afvejning mellem disse hensyn kan foretages ud fra viden om de konkrete test.

For et givent sæt af opgaver/opgavetyper er der en afvejning mellem, hvor mange forskellige dimensioner af fagområdet, der måles, hvor præcist hvert område måles, og hvor længe det tager at gennemføre opgaven. Andre af gruppens anbefalinger går på at mindske dette dilemma ved fx at prioritere opgavetyper, der giver høj sikkerhed, og ved at testen kan deles op over flere dage. Ikke desto mindre vil der skulle foretages en afvejning. I forhold til målesikkerhed ville det fx være at foretrække at have en *standard error of measurement* (SEM) på under 0,2, men det ville – alt andet lige – kræve nogle meget lange test. For at kunne afveje disse hensyn anbefales det, at der foretages estimer af disse forhold (målesikkerhed, fagområder og længde) for de konkrete opgaver/opgavetyper, som udvikles til fagene, og for de danske elever, som skal gennemføre testene.

Argumenter for at følge

- I. Selvom der også vil være usikkerhed på disse estimer (for målesikkerhed, faglig bredde og tidlig længde), som afhænger af opgavetype, vil det kunne give værdifuld viden, når disse hensyn skal afvejes.

Argumenter for ikke at følge

- I. Da sådanne data vil afhænge af opgavernes konkrete indhold, opgavetyper, opgavernes kvalitet, testtiden, der stilles til rådighed, fordelingen af sværhedsgraderne, selve testsituationen og en mængde yderligere faktorer, vil sådanne data kun vedrøre den konkrete test, som de er indhentet på, og derfor være uden værdi ud over testudviklingsfasen. I testudviklingsfasen vil der sædvanligvis blive set på målesikkerheden, og der vil blive arbejdet på at forbedre denne med forskellige greb. Det vurderes at være spild af tid og ressourcer at indlægge denne ekstra datagenerering, da den, som nævnt, ikke har bredere værdi. Hvis det var muligt at fremstille sådanne data med en mere generel nytteværdi, ville det være blevet gjort.

4.4. Testafholdelsen

ANBEFALING 58

Det anbefales, at lærere og elever efterfølgende kan genbesøge udvalgte testopgaver i forbindelse med det formative pædagogiske arbejde.

Når børnene og læreren konstaterer, at der er områder, der var udfordrende, kan det være en fordel at bruge tid på at knække koden, og derefter genbesøge testen.

Argumenter for at følge

- I. Eleven oplever succes, når de kan se, at de kan arbejde sig igennem en udfordring.

ANBEFALING 59

Det anbefales, at det er muligt for en elev at gå tilbage i testen og rette besvarelsen af tidligere løste opgaver.

Elever kan opleve det meget frustrerende, at de enten kommer til at gå videre i testen, inden de er færdige med besvarelsen af en opgave, eller at de lidt senere kommer i tanke om, at de kunne besvare en opgave på en anden måde, men ikke har mulighed for at gå tilbage og rette den pågældende opgave.

Argumenter for at følge

- I. Denne funktion øger muligheden for flere rigtigt besvarede opgaver.
- II. Denne funktion sikrer færre situationer, hvor elever frustreres over ikke at kunne rette en opgave, og giver således større tilfredshed med testen hos eleverne.

- III. Dette er en simpel forlængelse af principperne i en papir-blyant test, man har foran sig og vil støtte almindelig respondent adfærd.
- IV. Medvirker til at nedsætte elevernes følelse af et dårligt testforløb.

Argumenter for ikke at følge

- I. Dette må i så fald implementeres med nogle restriktioner, hvis det samtidig tillades, at det skal være muligt at stoppe og genoptage testen (jf. anbefaling 60).
- II. Kræver at der i senere opgaver ikke gives nogen oplysninger, der kan medvirke til besvarelsen af tidligere opgaver, hvilket ikke er usædvanligt at bygge ind i en evne eller færdighedstest.

ANBEFALING 60

Det anbefales, at det skal være muligt at stoppe og genoptage testen samme dag.

Børn kan koncentrere sig i forskellige tidslængder. I testsituationer er de meget ofte pressede af at præstere.

Argumenter for at følge

- I. Det kan øge muligheden for eleverne at holde koncentrationen.

Argumenter for ikke at følge

- I. Hvis fremgangsmåden følges, skal det præciseres, hvor mange gange testen kan stoppes, hvor længe m.v., da der ellers kan opstå en række problemstillinger, idet der ikke stilles ensartede testomstændigheder, og validiteten og/eller reliabilitet kan påvirkes.

ANBEFALING 61

Det anbefales, at man kan følge hastigheden på den enkelte elevs besvarelse på et eksternt *device* under afholdelsen af testen.

Det vil betyde, at læreren har mulighed for at opdage, hvis en elev enten er for hurtig eller er gået i stå i testen, og at læreren har mulighed for at understøtte en hensigtsmæssig testafvikling for den enkelte elev.

Argumenter for at følge

- I. Det letter støtten til de enkelte elever undervejs, at læreren eventuelt opdager, at en elev er for hurtig eller for langsom i gennemførelsen af testen.

ANBEFALING 62

Det anbefales, at læreren ved bookingen af testen mindes om at tage højde for elever med:

- **Særlige vilkår**
- **Hjælpe midler**
- **At eleverne er blevet instrueret i gennemførelse af testen.**

Lærerens hverdag er fuld af gøremål. Booking af test er et af dem, og det går måske så hurtigt, at læreren ikke er opmærksom på, at nogle elever har særlige behov.

Argumenter for at følge

- I. Det vil sikre, at elever med særlige behov tilgodeses.

- II. Det vil sikre, at alle elever er bekendt med, hvad det er for en test, de skal til at lave.
- III. Det vil sikre lige vilkår for alle elever i Danmark.

Argumenter for ikke at følge

- I. Det er vigtigt, at skolerne og lærerne får god og standardiseret information om gennemføringen af testerne, men et sådant krav vil medføre unødvendig bureaukratisering.

ANBEFALING 63

Det anbefales, at det sikres, at elever med dansk som andetsprog, fra ikke-middelklassebaggrund og fra ikke-urban baggrund, ikke stilles ringere alene på grund af deres baggrund.

Elever med dansk som andetsprog, fra ikke-middelklassebaggrund og fra ikke-urban baggrund, kan have andre erfaringer end elever med dansk som førstesprog, fra middelklassen og fra byerne, fx fra boligen, familielivet, omgivelserne, medierne osv. Det kan give sig udslag i, at opgaveindhold, der tager udgangspunkt i en "typisk" situation, kan være mindre velkendt for grupper af børn. Det er derfor nødvendigt at afprøve opgaver med tilstrækkeligt store grupper af børn med ikke-majoritetsbaggrund.

Argumenter for at følge

- I. Elever skal stilles lige i testsituationen.
- II. Med analyse af *Differential Item Functioning* (jf. anbefaling nr. 10) kan det testes om grupper af elever klarer sig uforholdsmæssigt dårligere på en given opgave.

Argumenter for ikke at følge

- I. Der savnes videnskabeligt belæg for, at elever fra "ikke-middelklassebaggrund" og "ikke-urban baggrund" skal sikres særligt, i forhold til ikke at stilles ringere alene på grund af deres baggrund (det antages, at der menes i forbindelse med Folkeskolens Nationale Færdighedstest). Se alternativ *anbefaling nr. 64*.

ANBEFALING 64

Det anbefales, at det sikres, at opgaveindholdet stiller elever lige, således at grupper af elever ikke forfordes i deres testresultater. Særligt bør der være fokus på, at elever stilles lige på tværs af grupper defineret af køn og sprog og andre relevante grupper.

Forskellige undergrupper af elever kan have forskellige erfaringer med det opgaveindhold, de præsenteres for i en opgave og opgaveindholdet kan derfor være mindre velkendt for nogle grupper af elever, eksempelvis grupper defineret ved sprog eller køn. Dette kan lede til systematisk forfordeling af grupper af elever, grundet den systematiske skævvridning forårsaget af, at nogle opgaver fungerer systematisk forskelligt for undergrupper af elever (jf. *differential item funktion* i *anbefaling nr. 10*). Dette bør naturligvis undgås, da alle elever har ret til et så retvisende resultat som muligt uafhængigt af deres sproglige, kønsmæssige eller anden relevant baggrundsgruppering.

Argumenter for at følge

- I. Elever skal stilles lige i testsituationen.

ANBEFALING 65

Det anbefales, at beslutningen om at tage en frivillig test tages så tæt på klasseværelset som muligt.

Med respekt for den rolle som spilles af kommunen og af skoleledelsen er formuleringen "tæt på klasseværelset" valgt for at signalere, at læreren bør spille en central rolle i beslutningen om at benytte frivillige test.

Argumenter for at følge

- I. Hvis begrebet frivillighed skal give mening, er det vigtigt, at de frivillige test også lokalt opleves som reelt frivillige.
- II. Det kan være en udfordring at koordinere hver enkelt af de frivillige test i hver enkelt fag med den pædagogiske tilrettelæggelse, som sker i den enkelte klasse. Derfor er det vigtigt, at tilvalget af den frivillige test sker tæt på denne proces.

Argumenter for ikke at følge

- I. En undtagelse til denne anbefaling kan være, i forbindelse med specifikke indsatser eller tiltag i en skole/kommune, hvor det er nødvendigt, at beslutning tages på højere (skole/kommunalt) niveau.

4.5. Øvrige temaer

ANBEFALING 66

Det anbefales at styrke udviklingen af Folkeskolens Nationale Færdighedstest med eksperimenter og følgeforskning i årene frem til skoleåret 2026/27.

Ekspirerter og følgeforskning kan sætte ind på områder som testenes design og varighed, de praktiske rammer for testenes afvikling, samt udvikling af dialogformer mellem de forskellige parter, der medvirker til styrkelse af den lokale evalueringskultur. Metoderne kan spænde fra kontrollerede designs til kvalitative studier.

Argumenter for at følge

- I. Det er svært at forudsige alle aspekter af Folkeskolens Nationale Færdighedstest og deres praktiske brug. Det er et komplekst projekt. Med en løbende afprøvning og vidensbaseret opfølgning vil man få den bedste udnyttelse af tiden frem til den fulde implementering af Folkeskolens Nationale Færdighedstest i skoleåret 2026/27.

Argumenter for ikke at følge

- I. Det kan blive omkostningskrævende.

ANBEFALING 67

Det anbefales årligt at indsamle læringspunkter vedrørende lærernes oplevelser af relevans og anvendelighed af de tilbagemeldinger, lærere modtager fra Folkeskolens Nationale Overgangstest, således at disse læringspunkter indgår i arbejdet med udviklingen af Folkeskolens Nationale Færdighedstest.

Selv om Folkeskolens Nationale Overgangstest i sagens natur er midlertidige, har de nogle fællestræk med de kommende Folkeskolens Nationale Færdighedstest. Det gælder fx det lineære testprincip. Ligeledes er det centralt, hvorvidt lærerne oplever formatet for visning af testresultater som overskueligt, forståeligt og omsætteligt. Man kan med fordel afprøve hensigtsmæssige tilbagemeldingsformater allerede i overgangsperioden – med lærernes vurdering af testenes relevans og anvendelighed som centralt pejlemærke. Læringspunkterne kan indsamles ved en afgrænset dataindsamling i form af et kort spørgeskema til en stikprøve og/eller kvalitative data.

Argumenter for at følge

- I. Jo flere sten, der kan ryddes af vejen inden introduktionen af Folkeskolens Nationale Færdighedstest i fuld skala, jo større er sandsynligheden for, at disse test vil fungere fra start.
- II. I modsætning til introduktionen af de nationale test i sin tid, findes der denne gang tid til mere grundig læring og afprøvning. Den tid bør udnyttes bedst muligt – med fokus på relevans og pædagogisk anvendelse.

Argumenter for ikke at følge

- I. Lærerne er i forvejen presset på tid og bør ikke få yderligere bureaukratiske opgaver.
- II. Da de nye Folkeskolens Nationale Færdighedstest formentlig ikke vil være sammenlignelige med Folkeskolens Nationale Overgangstest, hverken i indhold eller tilbagemeldingsformat, vurderes monitorering af Folkeskolens Nationale Overgangstest at være overflødig. I stedet bør der være en evaluering af de nye Folkeskolens Nationale Færdighedstest, når disse startes.

ANBEFALING 68

Det anbefales, at det sikres, at den enhed, fx Styrelsen for Undervisning og Kvalitet og Styrelsen for It og Læring, der har særskilt ansvar for Folkeskolens Nationale Færdighedstest, også har en kritisk mængde af eksperter i pædagogiske test. Desuden anbefales det, at der etableres mekanismer til fastholdelse af disse eksperter, således at Folkeskolens Nationale Færdighedstest-historikken bevares og kontinuiteten sikres.

Ekspertise i pædagogiske test kræver både indgående kendskab til det faglige område, testudvikling og psykometri, og der er ikke mange, der tilhører denne faglige kreds. Det er væsentligt, at enheden, som har ansvaret for Folkeskolens Nationale Færdighedstest, har en sådan ekspertise hos flere medarbejdere, som således ikke er generalister. Det er tillige vigtigt for historikken og kontinuiteten, at denne ekspertise fastholdes i enheden.

Argumenter for at følge

- I. Kan bevare historikken i Folkeskolens Nationale Færdighedstest, så man undgår at lave samme fejl to gange og i stedet bevæger sig fremad hele tiden og dermed sikrer kontinuiteten.
- II. Folkeskolens Nationale Færdighedstest er et komplekst redskab at udvikle, idet det dækker en række fag, mange klassetrin og har flere anvendelser. Med den rette ekspertise i den ansvarshavende enhed, vil opgaven kunne løses bedre – også fremadrettet.

ANBEFALING 69

Det anbefales fortsat at benytte sig af ekspertgruppens viden og kompetencer efter de samlede anbefalinger er afleveret til Børne- og Undervisningsministeriet.

Flere af anbefalingerne indeholder en meget stor kompleksitet og specialiseret viden. Derfor anbefales det, at relevante eksperter hentes ind til opgaver, hvor det kan være vigtigt at få uddybet og forklaret særlige emner og temaer, før der bliver truffet væsentlige beslutninger herom.

Særligt er ekspertgruppen opmærksomme på kompleksiteten i sammenhængene mellem de forskellige anbefalinger, hvor flere af anbefalingerne kan betragtes som indbyrdes forbundne kar - prioriteringer på et område kan medføre ændrede vilkår på andre områder.

Derfor stiller ekspertgruppen sig til rådighed, såfremt der er anbefalinger, der kræver ekstra uddybning eller forklaring. I det videre arbejde anbefales det endvidere, at den nuværende ekspertgruppe bliver repræsenteret i hvert af de kommende faglige arbejdsgrupper, således at viden videreføres.

Argumenter for at følge

- I. Det sikrer højere kvalitet i det videre arbejde.

Argumenter for ikke at følge

- I. Det er tidskrævende.

5 Et mindretal i ekspertgruppen anbefaler

5.1. Fagligt indhold

ANBEFALING 70

Det anbefales, at der udvikles flere forhåndsdeklarerede opgavesæt til brug i historie, der fx relaterer til forskellige kompetenceområder eller tager udgangspunkt i individuelle emner/kanonpunkter med henblik på, at eleverne kan aflægge test i stof, de har arbejdet med.

Det er politisk givet, at der skal være frivillige test i historie. Det er også givet, at undervisningen i historie i vidt omfang tilrettelægges lokalt. I kombination kan de to forhold føre til, at en frivillig test kan vedrøre emner, der ikke er gennemgået. Det kan medføre vilkårlige testresultater og uheldige testoplevelser.

For at sikre, at eleverne møder opgaver, der omhandler stof- og kompetenceområder, der har været arbejdet med i undervisningen på den pågældende skole, anbefales det, at der udvikles flere opgavesæt, hvor læreren udpeger et antal opgavesæt, som anvendes ved gennemførelse af testen.

Argumenter for at følge

- I. En forhåndsdeklarering er afgørende for at modvirke, at elever bliver testet i emner/færdigheder, som ikke er gennemgået i den enkelte klasse.
- II. Eleverne vil opleve at gennemføre en test, der tager udgangspunkt i et område af historiefaget, de kender.
- III. Forhåndsdeklarering og lærerens udvælgelse af en række opgavesæt vil i høj grad kunne modvirke *teaching to the test*.

Argumenter mod at følge

- I. Det risikeres, at der kommer et for stort fokus på netop det enkelte kanonpunkt som historiske begivenheder og ikke på at arbejde med generelle kompetencer fra historiefaget.
- II. Hvis der altid skal være flere test til rådighed at vælge imellem, skal der udarbejdes flere testopgaver.
- III. En udmelding om givne emner kan medføre en tendens til, at undervisningen tilrettelægges efter testemnerne. Denne effekt kan imidlertid dæmpes, hvis beslutningen om at tage testen reelt træffes frivilligt og lokalt, tæt på klasseværelset.
- IV. Der er tilsyneladende en modsætning i, at der skal testes i færdigheder i et fag, hvor der ikke indgår færdigheder. Det fremstår således som, at det ikke er muligt at inddrage historie som frivilligt testfag med en færdighedstest.
- V. Historie er et fag med meget få timer. Hvis der fra centralt hold sættes fokus på specifikke emner, vil lærerne i den korte tid der er, blot have tid til at forberede eleverne på de udpegede områder. Der vil ikke være tid til at tilpasse undervisningen til klassens specifikke forudsætninger og interesser.
- VI. Beslutning om afgrænsning af, hvad der skal testes, bør tages af dem, der udvikler testen, således at beslutningen kan ske på baggrund af grundige, faglige overvejelser, som udfoldes i et rammeværk for testen, og med afvejning i forhold til de øvrige rammer for testen.

5.2. Testresultaternes anvendelse

ANBEFALING 71

Det anbefales, at data kun udleveres til og kan tilgås af certificerede brugere.

Nationale test er et meget kompliceret redskab, som kan bruges forkert (og i meget høj grad er blevet brugt forkert). Det er afgørende, at brugerne forstår instrumentet og kender til korrekte brugsformer. Det anbefales derfor, at der udvikles en test, som brugere skal bestå for at få adgang til data. Brugere er embedsfolk i stat og kommune, forskere, konsulenter, skoleledere, faglige vejledere mv.

Argumenter for at følge

- I. Nationale test er så stærkt et redskab, at de bør anvendes med forsigtighed og på baggrund af viden om, hvad der testes, og hvordan resultaterne kan bruges og ikke kan bruges.
- II. Folkeskolens Nationale Færdighedstest tillægges meget stor betydning i såvel offentlighed som i forvaltninger og på skoler. Der tages beslutninger, som indimellem har betydelige konsekvenser, og det er derfor vigtigt, at de beslutninger ikke tages på et forkert grundlag.
- III. Forskning, der baserer sig på Folkeskolens Nationale Færdighedstest, skal som al anden brug ske på oplyst grundlag. Men det står forskere frit for at udvikle alternative metoder til anvendelse.

Argumenter for ikke at følge

- I. Ifølge den politiske aftale skal det fremtidige evaluerings- og bedømmelsessystem bidrage til mindre bureaukrati. En certificering af alle brugere vil være i modstrid med det.
- II. Der er en risiko for, at et krav om obligatorisk certificering af en bredt sammensat brugergruppe vil møde skepsis og derved mindske anvendelsen af testresultaterne. En sådan ordning bør derfor være et tilbud til brugerne om kvalificering.
- III. Spørgsmål om, hvordan testene skal forstås, og hvad de kan bruges til, skal være åbent for forskning og debat. Derfor skal der ikke være nogen, som på forhånd har retten til at definere en test, som andre skal bestå for at få adgang til at undersøge data.
- IV. Folkevalgtes adgang til data skal ikke begrænses af teknokraters bestemmelser af, hvordan testene skal forstås eller bruges, ligesom de folkevalgte skal have ret til at delegerer opgaven med at anvende data, som de ønsker, til deres embedsmænd.
- V. Brug af test er et omfattende begreb. Brugere inkluderer også elever, forældre og folkevalgte politikere i kommunerne. Det vil være udemokratisk og urealistisk at kræve særlige forhåndskundskaber og godkendelsesprocedurer for at kunne anvende test.

ANBEFALING 72

Det anbefales, at testresultater fra Folkeskolens Nationale Færdighedstest afrapporteres som antal rigtige svar inden for hvert målte område suppleret med en oversigt over typiske faglige udviklingsforløb.

Det er et væsentlig hensyn, at testresultaterne er letforståelige for alle brugere. Hvis testene er designet ud fra de psykometriske principper angivet i *anbefaling nr. 10*, vil antal rigtige svar (sumscore) have en tæt nok sammenhæng med andre opgørelser af resultatet (såsom en Rasch-score) til, at værdien af en letforståelig score opvejer hensyn til andre måder at afrapportere test scores.

5.3. Øvrige temaer

ANBEFALING 73

Det anbefales, at der etableres mulighed for, at skoler, der ønsker at benytte et andet certificeret testværktøj i et eller flere fag, som i al væsentlighed kan sidestilles med Folkeskolens Nationale Færdighedstest i forhold til testens faglige indhold, kan søge om dette gennem en simpel proces.

Argumenter for at følge

- I. Lokal praksis udviklet tæt på eleverne kan give en bedre testoplevelse.
- II. Testværktøjer, som lærere/faglige vejledere selv har udvalgt, giver større ejerskab.
- III. Folkeskolen frisættes i samme omfang som de private skoler til at træffe det valg, der lokalt vurderes at give den bedste evaluering af eleverne.
- IV. Det har været afprøvet med succes under Frihedsforsøget.

Argumenter mod at følge

- I. Der vil blive indsamlet færre data til brug for sammenligning af skoler og elevgruppe.
- II. Der bør ikke indføres flere test i folkeskolen, end der allerede er.
- III. En lokalt udviklet test, vil alt andet lige ikke kunne udvikles med sammen præcision og kvalitet som Folkeskolens Nationale Færdighedstest, medmindre der er meget store ressourcer til rådighed.
- IV. Da certificeringsordningen ikke er kendt, kan ekspertgruppen ikke forholde sig til dette.

ANBEFALING 74

Det anbefales, at alle landets skoler, også privat- og friskoler, skal gennemføre Folkeskolens Nationale Færdighedstest.

Argumenterne for at gennemføre obligatoriske Folkeskolens Nationale Færdighedstest på folkeskolerne gælder på samme vis for de øvrige skoletyper.

Argumenter for at følge

- I. Forskning mangler viden fra op mod en femtedel af elevpopulationen.
- II. Statslige myndigheder mangler indsigt i status for elever på privatskoler.
- III. Forældre, der er imod nationale test, kan fravælge folkeskoler alene af den grund.
- IV. Hvis ikke de private skoler tager Folkeskolens Nationale Færdighedstest, er det ikke muligt for beslutningstagerne at følge den faglige udvikling i Danmark som helhed.

Argumenter for ikke at følge

- I. Det bør være frivilligt for privat- og friskoler at deltage i Folkeskolens Nationale Færdighedstest, da det er vigtigt at værne om autonomien i den danske model for disse skoler.
- II. At det må være muligt, at man i Danmark driver skole med en anden didaktisk- og pædagogisk metode og derfor skal den danske folkeskole ikke ensrettes testmæssigt.
- III. Det er et gode i et demokratisk, pluralistisk samfund, at vi har flere forskellige skoleformer.
- IV. Forældre, der er imod Folkeskolens Nationale Færdighedstest, kan fravælge folkeskoler alene af den grund. Hvis en familie med åbne øjne på et velinformeret grundlag ønsker en skole uden nationale test, er det en fordel, at de har den valgmulighed.
- V. I det omfang, som familier måtte fravælge folkeskolen af denne grund, kan det tjene som påmindelse om, at folkeskolen ikke bør fremstå som alt for domineret af en test- og præstationskultur.

- VI. Stigende mistrivsel funderet i præstationsorienterede kvalitetssikringsredskaber, skal ikke overføres til fri- og privatskoler. Det bør i stedet være frivilligt for alle skoler, lærere og elever, om de vil deltage i Folkeskolens Nationale Færdighedstest.
- VII. Med den stigende mistrivsel blandt folkeskoleelever i Danmark, bør en testpraksis som nationale test, der konkret har vist at have en mistrivsel fremmende effekt (Bundsgaard & Puck 2016³⁵), ikke udbredes til elever i andre skoleformer, da dette ville bidrage yderligere til mistrivsel blandt danske børn og unge.

ANBEFALING 75

Det anbefales at supplere Folkeskolens Nationale Færdighedstest med et mindre antal opgaver, fx 5-10 procent, der rettes af enten klassens lærer, et korps af specielt uddannende testrettere eller et korps af censorer.

Med de teknologiske muligheder, der pt. er til rådighed, er der dele af fagene, som er vanskelige at teste ved brug af digitale selvrettende test. Dette krav begrænser de typer af opgaver, som kan stilles og de svar elever kan give – og dermed hvilke kompetenceområder, testen kan teste. I store internationale undersøgelser, som fx PISA, rettes en mindre del af opgaverne af et korps af specielt uddannede testrettere. Det anbefales derfor at supplere Folkeskolens Nationale Færdighedstest med et mindre antal opgaver, der ikke er digitalt selvrettende og at undersøge omkostningerne i forhold til tre scenarier – opgaverne rettes af: 1) klassens lærer, 2) et korps af specielt uddannende testrettere, som ved PISA, og 3) et korps af censorer som ved folkeskolens prøver. Fordelen ved det første scenarie er, at læreren får indblik i egne elevers viden, færdigheder og kompetencer, mens fordelene ved de to øvrige scenarier er, at der kan laves nationale opgørelser af elevernes besvarelser af de supplerende opgaver. Uanset det valgte scenarie, skal testudviklerne – eventuelt i samarbejde med testrettere eller censorer – udarbejde overordnede kategorier af elevernes besvarelser af disse opgaver og foreslå ideer til opfølgning.

Argumenter for at følge

- I. Følges anbefalingen kan elever besvare de supplerende opgaver på andre måder end de øvrige testopgaver, fx ved at forklare og at begrunde. Det giver muligheder for at få et større indblik i elevernes faglighed, og at udvikle tilbagemeldinger til lærere (og andre testbrugere), som er af højere kvalitet og dermed af større formativ værdi for udvikling af undervisning.
- II. Supplementet behøver ikke at forøge udgifterne til Folkeskolens Nationale Færdighedstest. Merudgiften ved at aflønne lærere eller at uddanne og vedligeholde et korps af testrettere eller censorer kan formodentlig opvejes af færre udgifter til at udvikle selve testen.
- III. Hvis læreren retter en del af testen manuelt, kan det være incitament for læreren til at fordybe sig i elevernes besvarelser af testen.

Argumenter for ikke at følge

- I. Dette vil medføre øget brug af ressourcer, og det vil forhindre hurtig tilbagemelding til eleverne og lærerne.
- II. Er forbundet med relativt store merudgifter til at specialuddanne rettere (kontinuert), undersøgelser af inter-rater reliabilitet osv.
- III. Den objektive bedømmelse mistes.

³⁵ Bundsgaard J. & Puck M. R. (2016). *Nationale test: Danske lærere og skolelederes brug, holdninger og viden*. DPU, Aarhus Universitet. Center for Anvendt Skoleforskning ved University College Lillebælt: <https://dpu.au.dk/fileadmin/edu/Forskning/Ebog - Nationale test - FINAL - september 2016.pdf>

6 Bilag

Der er udarbejdet to positionspapirer, her listet alfabetisk efter forfatternavn:

- Mads Poulsen
- Rolf Vegar Olsen, Tine Nielsen, Jeppe Bundsgaard, Mads Poulsen og Simon Calmar Andersen

Disse positionspapirer er inkluderet som bilag til ekspertgruppens anbefalinger.

Positionspapir om placering af ansvar for opfølgning hos lærere

Forfatter: Mads Poulsen

Der er givet en række adskilte anbefalinger om, hvordan der skal gives tilbagemeldinger. Uanset hvilke anbefalinger der vælges, opfordres der til, at det væsentlige ansvar for at udstikke retningen for den enkelte elev og klassen tydeligt placeres lokalt hos læreren og skolens fagpersonale. Dette kan være i konflikt med andre ønsker. Fx kan ønsker om at øge testens pædagogiske anvendelighed anspore til at lade testen give karakteristikker af elevers udviklingsmuligheder. Men hvis anvisninger gives i en automatiseret tilbagemelding fra testen, så begrænser det lærerens faglige råderum for at tilrettelægge opfølgning og undervisning. Det gælder især, hvis anvisningerne gives direkte til elever og forældre. Det er også muligt, at læreren tilsidesætter sin egen samlede vurdering af eleven, hvis der gives konkrete anvisninger.

Det er stadig vigtigt, at forældrene får en tilbagemelding, men den bør have generel karakter, som det er op til læreren at konkretisere. Det er oplagt, at der udarbejdes støtte til lærerens arbejde med at konkretisere, men det væsentlige er, at det er lærerens opgave at anvise retning over for elever og forældre.

Man kunne have forhåbninger om, at FNF faktisk ville kunne give bedre anvisninger end den gennemsnitlige lærer. Det er underskrivernes vurdering, at det bestemt ikke kan garanteres, bl.a. fordi vi ikke har set eksempler på, at et sådant omfattende system er udviklet før, hverken i Danmark eller internationalt. Det er komplekst at anvise om individuel opfølgning, og der er begrænset forskning at basere anvisninger på. Anvisninger vil derfor let få karakter af idiosynkratiske holdninger. Når automatiserede anvisninger ikke kan evidensbelægges solidt, så er det bedre at overlade ansvaret til det lokale fagpersonale, der kan tage hele elevens situation, inklusive testresultater, i betragtning.

Positionsnøtat

Rapporten præsenterer en række anbefalinger for de nye færdighedstest. Meningen med udvalgets arbejde har netop været at formulere et sådant sæt af enkeltstående anbefalinger, som kan vurderes og tages stilling til en efter en. Dette har forenklet og gjort det muligt for gruppen at have et overblik og gjort det muligt at gennemføre på kort tid.

Det er imidlertid vigtigt at påpege at selv om man kan stille sig bag de enkelte anbefalinger, så er det samtidig umuligt at se for sig at alle anbefalingerne kan blive fulgt når testene faktisk skal udvikles og implementeres. Anbefalingerne repræsenterer netop ikke en samlet og kohærent beskrivelse af et rammeværk som beskriver hvilke formål og kvalitetskriterier testene skal efterleve.

De miljøer som i sidste ende skal udvikle testene er afhængige af at etablere et samlet og indre konsistent rammeværk før de kan lave gode, meningsfulde og brugbare test. Vi ser for os at i processerne som skal igangsættes i bagkant af udvalgets arbejde, så bør der etableres to typer rammeværk med forskellige funktioner. For det første bør der etableres et rammeværk med nogle centrale og generelle principper samt guidelines, som gælder for alle testene som skal udvikles. I tillæg bør der udvikles flere specifikke rammeværk som beskriver det faglige indhold og eventuelt fagspecifikke principper for de enkelte test.

I dette arbejde vil man uundgåelig være nødt til at træffe en række valg ganske tidligt i processen. Det vil også ganske hurtigt i processen blive tydeligt at, der er flere hensyn som må opvejes mod hinanden. Vi ønsker derfor at påpege at nogle egenskaber ved de nye test bør prioriteres højere end andre hensyn. Flere af anbefalingerne omtaler også momenter som på ulige vis berører disse prioriteringer:

- Udgangspunktet for arbejdet med de nye færdighedstest er at de skal være brugbare værktøjer for lærerne i forhold til behov at give nyttig information, som kan anvendes formativt og i tilrettelæggelsen af undervisning i klasseværelset og for den enkelte elev, og i tillæg at de kan give anvendelig information for det lokale arbejdet med kvalitetsudvikling i den enkelte skole.
- En test som ikke tilfredsstillende de helt grundlæggende krav til psykometrisk kvalitet giver begrænset og usikker information om enkeltelever. Principper som identificerer standarder for henholdsvis målefejl og reliabilitet skal derfor prioriteres højere end andre hensyn.
- En test kan ikke favne det hele og alt som det er vigtigt at vide noget om. Derfor skal de endelige test måle skolefaglige størrelser som alment opfattes som at være helt centrale forudsætninger for videre læring i faget. (I praksis er informationen fra test begrænset af en række forhold, så som længde/varighed av testen, opgaveformater mm.)

Rolf Vegar Olsen
Tine Nielsen
Jeppe Bundsgaard
Mads Poulsen
Simon Calmar Andersen

