

resultatet representativt for det faktiske ferdighetsnivået, mener de. Drillingen blir en form for akademiske steroider som gir hovedstadens skoler kunstig overdimensjonerte testresultater.

Disse lærerne hevder at Oslo-skolen jukser ved nasjonale prøver.

JUKS ELLER IKKE JUKS? Har de kritiske lærerne rett i at det som foregår er juks? Eller er metodene i Oslo-skolen snarere en mer effektiv undervisningsform? Her står det en fundamental faglig uenighet innen hovedstadens skolesystem. Den kommer til å prege hele Undervisnings-Norge i årene som kommer.

Når lærere klager på såkalt drilling av elevene, er det nærliggende å spørre: Hva er egentlig problemet?

«Drilling» er jo ikke noe annet enn målrettet innlæring av det elevene faktisk skal lære. Økt satsing på grunnleggende ferdigheter er vedtatt av Stortinget. Grunnleggende ferdigheter er nettopp hva nasjonale prøver skal måle. Også de er innført av Stortinget. Så hva er problemet med at skolene intensiverer øvingen på det som måles, når disse basisferdighetene faktisk er noe av det viktigste skolen skal lære barna våre?

Det eksisterer ikke noe problem, mener Oslos skolebyråd Torger Ødegaard: «Han sier han ikke ser noe galt i at elevene setter av et halvår til å terpe på de kommende nasjonale prøvene til høsten,» ifølge *NTB* 23. mai 2011.

«Det må da være greit at elevene bruker tid på å trene på lesing, regning og grunnleggende ferdigheter. Idrettsfolk trener vel også før de skal på banen, og det man trener på, blir man jo god i,» sier Ødegaard.¹¹

Er det fornuftig å se på nasjonale prøver som en idrettsgren elevene skal trene seg opp i? Her må vi ta et skritt tilbake og se på hva nasjonale prøver egentlig *er*. Vi bør vite noe om prestasjonstestingens *rolle* i et system for måling og kvalitetsutvikling

i skolen. Her skal vi importere litt akademisk bistand fra skoleprestatjonstestingens fremste foregangsland.

ET STØRRE KUNNSKAPSDOMENE. Ingen land er så ivrige på prestasjonstesting innen det offentlige skolesystemet som USA. Få har forsket mer på området enn Daniel M. Koretz. Koretz er professor ved Harvard Graduate School of Education, medlem av USAs National Academy of Education og leder forsker-nettverket «International Project for the Study of Educational Accountability». I 2008 publiserte han en bok som handler om nettopp de standardiserte testenes styrker og begrensninger som redskap for skoleutvikling.¹²

I boken *Measuring Up* påpeker Koretz den utbredte misforståelsen at prestasjonstester i skolen måler *det som elevene skal kunne*. Det gjør de ikke.

En prøve på 90 minutter kan ikke kartlegge hele elevens ferdighet, eller mangel på sådan, verken i lesing, regning eller engelsk. Prøven må nødvendigvis være en stikkprøve. Derfor halter byråd Ødegaards sammenligning med en idrettskonkurranse. Når løpere konkurrerer i 100 meter sprint, er det fordi vi vil vite hvor gode de er til å løpe 100 meter sprint. Når elever tar nasjonale prøver i regning, er det *ikke* fordi vi vil vite hvor gode de er i nasjonale prøver i regning. Det som interesserer oss, er hvor gode de er i *regning*, som er et langt større «domene» av kunnskap og ferdigheter.

Det er ikke praktisk gjennomførbart å kartlegge *alle* aspekter ved ferdighetene «lesing» og «regning», og særlig ikke i løpet av 90 minutter. Oppgavene i prestasjonstesten utgjør derfor en liten del av noe mye større, som vi egentlig er ute etter: Elevens mestring innen et større «kunnskapsdomene» på det området som testes. Det lille utvalget av ferdigheter som testes, må «representere» den større helheten.

Det betyr at testresultatene er verdifulle hvis, og bare hvis, de forteller oss noe om elevens ferdigheter innen det større domenet. Det er med andre ord ikke slik at nasjonale prøver «avleser» elevens mestringsnivå. Vi må foreta en *slutning* fra testresultatet (utvalget) til elevens mestringsnivå (helheten). Den typiske slutningen er at elevens testprestasjoner forteller oss hvordan eleven *ville* ha prestert innen hele domenet, hvis vi kunne testet dét. En slik forbindelse mellom testresultat og elevens faktiske ferdigheter forutsetter at resultatet er representativt.

Hvis en prestasjonstest i skolen skal være representativ, kan den ikke bestå av tilfeldig valgte oppgaver basert på pensum. En treffsikker test, skriver Daniel Koretz, må bygge på et nøye gjennomtenkt utvalg av de konkrete kunnskapene og ferdighetene som skal testes.

Når vi skal teste ferdighetene i engelsk med nasjonale prøver i åttende klasse, må vi for det første vite hva som er forventet kompetanse, ifølge læreplanen. Ett kompetansemål her er for eksempel at eleven skal kunne «bruke grunnleggende regler og mønstre for uttale, intonasjon, rettskriving, grammatikk og ulike setningstyper». ¹³ Så må vi bestemme hvilke *aspekter* av målene som er viktige å teste (hvilke deler av engelsk grammatikk, for eksempel) og hvilken vekt som skal legges på hvert enkelt aspekt. Slik utvikler vi et ganske detaljert kart over hva vi ønsker å undersøke. Dette kartet danner *utgangspunktet* for å lage prøvene.

Prøvene skal sortere elevene på ulike mestringsnivå. Derfor må vi ta rede på hva som skiller gode elever fra middels og middels elever fra dårlige, det vil i barneskolen si de 25 prosent øverste, de 50 prosent på midten og de 25 prosent nederste. Utkast til nasjonale prøver blir derfor prøvekjørt på et utvalg av skoler. Ved å analysere svarene fra disse elevene kan testmarkere finne hva som er typiske kjennetegn ved gode, middels og dårlige prestasjoner. Siktemålet med utvalget av oppgaver

er ikke primært å kartlegge nøyaktig hva den enkelte elev er god og dårlig til, men å finne begreper og oppgaver som effektivt «avslører» om eleven skal plasseres i de øverste 25, midterste 50 eller nederste 25 prosent av fordelingen.

Derfor er det ikke tilfeldig hvilke utfordringer elever møter ved nasjonale prøver. Dermed er det heller ikke umulig for skolene å drille elevenes ferdigheter i riktig retning.

Hva sier Koretz om det store stridsspørsmålet: Er det greit å drive med prøvedrilling à la Oslo-skolen, eller er det en form for juks? Han illustrerer med et eksempel.

Tenk deg at du skal starte et tidsskrift og søker etter medarbeidere. Du mottar en masse søknader fra universitetsstudenter. Hvem skal du ansette? Du bestemmer deg for å velge på grunnlag av hvor solid vokabular søkerne har. Du bestemmer deg for å teste 40 ord som er egnet til å skille den språkkompetente søker fra den som mangler riktig type vokabular. 40 ord er en dråpe i havet når man skal lage tidsskrift. Testen er verdifull bare hvis utvalget av oppgaver representerer det vi skal teste – søkerens vokabular. Ifølge Daniel Koretz er dette «utvalgsprinsippet» det mest grunnleggende innen prestasjonstesting i skolen. Det er ikke de utvalgte oppgavene i seg selv, men det at vi kan anta at elevenes prestasjoner på dem er representative for elevenes generelle nivå, som gjør at prestasjonstester kan være verdifulle. Her nærmer vi oss årsaken til at drillingen av Oslos skoleelever kan være problematisk.

Sett at jeg får tak i lista med 40 ord. Sett at jeg begynner å tilby alle jobbsøkerne et lynkurs i tidsskriftarbeid. Der lærer jeg dem utelukkende betydningen av akkurat disse 40 ordene. Så tar de prestasjonstesten din.

Testen er nå nærmest verdiløs. Full pott på prøven trenger ikke bety noe som helst ut over at søkeren mestrer 40 middels

vanskelige ord. Hadde testen vært gjennomført med tegnspråk, kunne man i prinsippet drille apen Julius til å demonstrere solid vokabular. Du risikerer å starte tidsskrift med et meget språksvakt individ som din redaksjonssekretær.

Vi bør legge merke til at problemet *ikke* er at «testen er dårlig». Det er den ikke. Problemet er heller ikke at testen «ikke måler alt som er viktig», for det er ikke det prestasjonstester gjør. Problemet er at *forbindelsen* mellom testresultatet og det som skal testes, er ødelagt. Forbindelsen er ødelagt fordi utvalget av oppgaver som testes, ikke lenger er representativt. Etter min drilløvelse er mestringen av de 40 ordene tvert imot helt urepresentativ for søkerens generelle nivå.

Her kan det innvendes at nasjonale prøver i Norge slett ikke består av 40 ord elevene kan pugge. Skolene vet heller ikke på forhånd hvordan årets prøver vil se ut. Det stemmer. Men skolene vet ganske mye.

Skolene har begynt å kartlegge hva slags oppgaver som typisk opptrer i de ulike prestasjonstestene. Hvis det aldri er geometri på regneprøvene, vil skolens resultater tjene på at lærerne legger vekk geometrikapitlet, i alle fall fram til elevene har gjennomført testen sin. En far med barn i Oslo-skolen forteller at dette skjedde i klassen til sønnen hans.

Hvis det er mange oppgaver med avkryssingsalternativ («multiple choice»), kan elevene læres opp i hvordan de øker sjansen for å skåre riktig uten å kaste bort verdifull tid på å stå fast og ønske: Eleven må eliminere de usannsynlige alternativene først, gjette blindt mellom de som står igjen, og komme seg videre på testen. Den gode nyheten er at ferdigheter innen *multiple choice* kan øke skolens gjennomsnittlige poengsum. Den dårlige er at få av utfordringene elevene senere skal møte i arbeidslivet, vil opptre i form av *multiple choice*.

Forberedelsen av elevene kan gjøres mer spesifikk enn som så. I januar 2011 begynte de østlige Oslo-skolene i området fra Alna til Stovner (skolegruppe C) hardkjøret fram mot nasjonale prøver påfølgende høst. Lærere på sjuende trinn ble kalt inn til et stort møte med områdedirektøren og skolelederne i området. Her fikk de presentert et ganske omfattende treningsopplegg.

Hvem er områdedirektøren? Som *Dagens Næringsliv* skriver, er Oslo-skolen organisert som «et konsern der elevenes resultater tilsvarer bunnlinjen». Rektorene har resultatansvar og lederavtaler, mens utdanningsdirektøren er «konsernsjef». Mellom disse to leddene finnes de viktige områdedirektørene.

Konsernet er nemlig delt opp i divisjoner, som er 20–25 skoler i samme geografiske område, og områdene har hver sin områdedirektør. Denne direktøren følger opp hver enkelt skoleledelse slik at den lokale skolen leverer sitt forventede bidrag til Oslo-effekten. Nå hadde altså direktøren for skolegruppe C kalt inn til møte for å få gjort noe kraftfullt med resultatoppnåelsen i området.

– Vi fikk et flammende innlegg fra områdedirektør Bjørn Indrelid, som skulle motivere til innsats fram mot høsten, forteller lærer Anne Jørgensen, som deltok fra Stig barneskole.

Direktøren presenterte et felles øvingsopplegg for alle skolene i området. Barnas prestasjoner skulle trenes opp i tre omganger gjennom våren.

– Vi skulle gjennomføre tre runder med alle tre prøver, både lesing, engelsk og matte. Barna skulle testes i februar, så i april og så igjen nest siste uke før ferien, forteller Jørgensen.

Det betyr ni prøver à 90 minutter, som alle skal forberedes, avholdes, rettes og dokumenteres.

– Vi oppfattet budskapet slik at vi skulle legge vekk annet arbeid i perioden fram til slutten av sjuende klasse, for å øve til nasjonale prøver.

Da NTB fortalte om dette treningsopplegget i nyhetssaken «Læreropprør mot konstant prøvepress», avviste områdedirektøren at dette var et pålegg fra hans side: «Jeg har aldri pålagt lærerne å gå bort fra sine planer eller å bruke våren til å øve på gamle prøvesett, men det er alle læreres ansvar å trene elevene i lesing og regning.»¹⁴

I forbindelse med dette «kick off»-møtet, som det ble kalt, sendte områdedirektøren ut en liste med «viktige ord» som burde øves inn før nasjonale prøver. Dette er ord som staben ved en av skolene har kartlagt som viktige ved nasjonale prøver på åttende trinn.

I e-posten står det:

Arbeidet må prioriteres sterkt veldig raskt. [...]

Viktige ord for hver uke i norsk / engelsk som brukes av alle lærere i hver time. [...] Dersom dere ikke er der ennå, kommer forslag til ord her slik at alle 7. trinns – rom har dette som store oppslag fra og med mandag:

Eng: fridge, actor, racehorse, rowboat, trained, advertisement, technology, changes, adult, cooking

No: varierer, argumenter, prosentandel, utsagn, skjønnlitteratur, kåseri, ekstrem, teknologi, energiformer, energiforbruk

Disse ordene er nødvendige å beherske for å ha sjansen for å svare riktig på nasj prøve i no / regning

Spørsmålet er hvor pålitelig styringsinformasjon de nasjonale prøvene kan gi, når skolene begynner å opptre som om dette

ikke er en stikkprøve, men en idrettsturnering. Noen av resultatene vil være kunstig overdimensjonerte ved hjelp av akademiske steroider av ymse slag, mens andre skoler ikke bruker slike metoder. I det første tilfellet vil ikke testen gi god styringsinformasjon, fordi utvalget av oppgaver ikke lenger er representativt.

Problemet med et ikke-representativt utvalg er lett å forstå når det ikke er snakk om et utvalg oppgaver, men mennesker som svarer i en politisk meningsmåling. Man får ikke en god prognose for det *generelle* valgresultatet ved å gjennomføre en valgdagsmåling inne i festsalen på Høyres Hus. På samme måte får man ikke en god prognose for skolens *generelle* kvalitet ved å sjekke akkurat de ferdighetene som skolen har drillt elevene i over lang tid.

I den nye drillskolen til Torger Ødegaard kan verken foreldre eller folkevalgte vite hva som er kunnskapsløft og hva som er kunnskapsbløff. Problemet kan illustreres med et eksempel fra IT-bransjen. Der testes mikrochips for å sette en tallverdi på hastigheten til nye datamaskiner. Slike tall brukes iherdig i markedsføringen, og de påvirker beslutningene når store virksomheter fornyer datamaskinparken. Men testene kan bare måle et utvalg av «ferdigheter» hos mikrochipen. Derfor har produsenter av mikrochips tilpasset «testoppgavene» til den bestemte chipen, eller omvendt, laget chips som presterer særlig godt på akkurat de operasjonene som testes. Resultatet blir mikrochips som viser høyere hastighet når de testes enn datamaskinbrukeren vil oppleve i det virkelige liv, når chipen er satt inn i sin sammenheng.¹⁵

Det er *anvendelsen* av de innlærte ferdighetene i praksis som er sannhetens øyeblikk for datamaskinen, ikke den enkeltstående målingen av isolerte operasjoner. Det samme kan sies om skolen. I begge tilfeller fører innretning spesielt mot testen til at testen mister sin verdi: Det blir umulig å trekke en gyldig

slutning fra testresultatet til hvordan ferdighetene vil fungere i virkeligheten.

Faktisk er problemet med nasjonale prøver i Norge enda mer alvorlig. For de brukes ikke bare til å dra slutninger om elevens nivå innen ferdigheten som testes, men som indikator på skolenes *generelle* kvalitet. Først generaliseres det fra testresultatet til elevenes generelle kompetanse, for eksempel i engelsk, så generaliseres det fra dette til kvaliteten på læreren, og deretter fra denne engelsklærers kvalitet til skolens generelle nivå, det vil si *de andre* lærernes kvalitet.

Disse ugyldige slutningene blir ikke mer holdbare av at skolene i Oslo gjennomfører en målrettet drilling fram mot nasjonale prøver. Snarere tvert imot. Når skolebyråd Torger Ødegaard stolt forsvaret sin oppfatning av nasjonale prøver som en slags idrettsturneringer, tyder det på at Høyre bygger sin politikk på en misforståelse av prestasjonstestens rolle i måling og utvikling av skolekvalitet.

– **OSLO-SKOLEN JUKSER.** Mens resultatfasaden skinner, foregår fabrikkeringen av elevprestasjoner ved hjelp av metoder som underminerer måleverktøyet. Dette er bare ett av problemene med Høyre og Frps forsøk på resultatmålstyrt kunnskaps-skole, ifølge lærerne ved Østskolen. Det som gjør dem mest sinte, er opplevelsen av å reduseres til fabrikkfunksjonærer i en fasadeskole der det eneste som teller, er det som telles.

– Alt handler om å pusse fasaden og se bra ut for dem lenger oppe i systemet, forteller «Svein» (40). Han gir et eksempel fra barneskolen hvor hans egen sønn går.

– De hadde en kartleggingsprøve i første klasse. Vi oppdaget at de drillet ungene i forkant på de vanskelige ordene som kom på prøven. Hva slags *kartlegging* er det?

Svein har vært lærer i rundt ti år. Jeg intervjuer ham i et