



Om at lave læseprøver

HOLGER JUUL, LEKTOR, PH.D., KØBENHAVNS
UNIVERSITET OG LENE MØLLER, TIDLIGERE
LEKTOR VED PROFESSIONSHØJSKOLEN
METROPOL

Testserien *Skriftsproglic udvikling* begyndte at udkomme i august 2009 og fylder således snart ti år. De første læseprøver i serien var trykte prøver med elever i 1. til 6. klasse som primær målgruppe. I 2015 kom der digitale udgaver til. Senest er der udkommet prøver til elever i 6. til 8. klasse (udelukkende i digitalt format). I artiklen fortæller vi om nogle af de overvejelser, vi som forfattere har haft om, hvordan prøverne skulle udformes.

Prøveserien *Skriftsproglic udvikling* er udformet af denne artikels forfattere og udgives af Hogrefe Psykologisk Forlag. Serien dækker 0. til 8. klassetrin og omfatter foruden prøver af bogstavkendskab og stavning disse læseprøver:

Ordlæsning

Ordlæseprøve 1-2 (1.-5. klasse)
Ordgenkendelse (6.-8. klasse)

Sætningslæsning

Sætningslæseprøve 1-2 (1.-5. klasse)
Sætningslæsehastighed (6.-8. klasse)

Tekstlæsning

Tekstlæseprøve 1-8 (1.-7. klasse)
Tekstforståelse Udsagn A og B (to parallelle prøver; 6.-8. klasse)
Tekstforståelse Cloze (6.-8. klasse)

Da et testforlag spurgte os, om vi ville udforme en ny serie læseprøver, satte det mange overvejelser i gang. Noget af det første, vi blev enige om, var at prioritere den praktiske anvendelighed højt. Prøveaktiviteter er ikke noget formål i sig selv, så generelt er vores holdning, at prøver skal være så kortvarige som muligt. Men samtidig skal resultaterne selvfølgelig give informationer, der er væsentlige for lærerens tilrettelægning af under-

visningen, hvilket kræver, at antallet af opgaver er stort nok til at give et nuanceret indtryk af hver enkelt elevs færdigheder.

Prøveresultater indebærer nærmest pr. definition en reduktion [...] men i mange sammenhænge er det faktisk en fordel.

Uanset hvor lange prøver man udsætter sine elever for, vil der være aspekter, der ikke kommer med. Prøveresultater indebærer nærmest pr. definition en reduktion. Det lyder jo ikke rart, men i mange sammenhænge er det faktisk en fordel, for formålet med prøver er oftere at få overblik over det væsentligste end at dokumentere verden i al dens kompleksitet. Udfordringen er at udforme prøver, så det, der træder frem, er noget væsentligt – uden at der går for meget tid fra den egentlige undervisning.

Tidsgrænser

Et af vores vigtigste valg var, at de fleste prøver skulle afvikles med ret korte tidsgrænser, men til gengæld have så mange opgaver, at det var svært for eleverne at nå igennem dem alle. Meningen var, at de bare skulle nå så meget, de kunne. Det har ganske vist den ulempe, at man ikke kan sammenligne eleverne på samtlige enkeltopgaver – men til gengæld ved man, at eleverne har arbejdet lige længe, og man undgår, at prøvesituationen bliver en klassekonkurrence om at blive først færdig. En anden fordel er, at prøverne er følsomme for et bredt spektrum af færdighedsniveauer – fra elever, der ikke når så mange opgaver eller laver mange fejl, og til elever, der får en høj score, fordi de både svarer hurtigt og laver få fejl.

Ord- og tekstlæsning

Læsning er mange ting. At læse sammenhængende tekster stiller selvsagt flere krav end at læse enkeltord (Elbro, 2014). Når det gælder begynderne, er det dog naturligt at fokusere på, om eleverne er i stand til at læse enkeltord, for hvis de har problemer med dét, er det der, man bør fokusere i undervisningen. Omvendt er hovedinteressen fra mellemtrinnet og frem, om eleverne kan læse tekster med god forståelse. Med *Skriftsproglig udvikling* har vi imidlertid valgt at tilbyde prøver af både ord- og tekstlæsning lige fra 1. til 8. klasse.

Selv blandt ældre elever kan det være relevant at afdække niveauet mht. ordlæsning; hvis elevernes læseforståelse er begrænset, kan usikker eller langsom ordlæsning være en vigtig del af forklaringen (Petersen & Rønberg, 2011). Og når det gælder elever i indskoling, kan man som lærer selvfølgelig *også* være interesseret i at vide, om eleverne, når de først har knækket koden, er i stand til at læse (simple) tekster med udbytte. Vi har på den baggrund valgt at lade det være op til læreren/skolen, hvad der skal fokuseres på. Man kan vælge en prøve af ordlæsning eller en prøve af tekstlæsning, alt efter hvad man ønsker belyst.

Hvad angår udformningen af tekstprøver, er der principielt to veje at gå. Man kan (som vi gør i prøverne *Tekstforståelse Udsagn* og *Tekstforståelse Cloze*) præsentere eleverne for en række relativt korte tekster med indlagte opgaver. På den måde kan de inden for prøvetiden møde eksempler på vidt forskellige emner og teksttyper, både fortællende og informerende. Den anden vej (som følges i *Tekstlæseprøve 1-8*) er at lade eleverne læse en længere afsluttet tekst. Fordelen her er, at læseprøven svarer godt til læsning i den virkelige verden, hvor man jo ofte beskæftiger sig med hele tekster. I forhold til afsluttede tekster kan anvendelse af tidsgrænser dog være uhenigtsmæssig, for så bliver langsomme læsere jo snydt for slutningen. De to nævnte tekstprøveformater (læsning af udvalgte passager vs. længere afsluttede tekster) har begge deres berettigelse, men lægger vægten lidt forskelligt: Vil man have bredde i teksttyper, eller vil man have en mere autentisk læsesituation?

Sætningslæsning

Som en mellemting mellem ord- og tekstlæseprøver kan man også vælge at bruge prøver af sætningslæsning. Sætningslæsning kræver mere af eleverne end prøver af ordlæsning, og samtidig er de hurtigere at gennemføre end prøverne af tekstlæsning. Desuden ser vi det som en fordel ved sætningsprøver, at de kan give et klarere indtryk af elevernes grundlæggende læsehastighed end tekstprøver. Ved tekstlæsning kan tænksomme elever bruge ganske lang tid på at reflektere over tekstens indhold, hvilket på mange måder er en god ting – men med den konsekvens, at de fremstår som langsomme læsere. I sætningslæseprøver er der normalt ikke helt så meget at reflektere over. I vores sætningsprøver gælder det bare om at nå at læse så mange sætninger som muligt inden for tidsgrænsen og at besvare de tilhørende opgaver uden fejl.

Overordnede scorer og delscorer

Uanset hvad læseprøver fokuserer på, bør de efter vores opfattelse munde ud i en samlet, overordnet score for hver elev, sådan at man som lærer får overblik. Det er imidlertid vigtigt at være opmærksom på, at den overordnede score i prøver på tid både påvirkes af præcisionen (hvor stor en del af opgaverne eleven har besvaret uden fejl) og af hastigheden (hvor mange opgaver eleven har besvaret inden for tidsgrænsen). Når man som lærer skal planlægge læseundervisning, er begge aspekter, præcisionen og hastigheden, ofte vigtige at få belyst

Uanset hvad læseprøver fokuserer på, bør de efter vores opfattelse munde ud i en samlet, overordnet score for hver elev, sådan at man som lærer får overblik.

Dermed bliver der imidlertid hele tre læsescorer for læreren at forholde sig til for hver elev. Vores valg har været at inkludere alle tre i de klasse- og elevrapporter, som man kan udtrække (via nettet), når en prøve er afsluttet. Den overordnede

score giver et indtryk af elevens generelle niveau, mens scorerne vedr. præcision og hastighed ofte giver et indtryk af, hvad eleven har brug for at arbejde med. Så længe læsningen er usikker, er det der, det pædagogiske fokus må lægges. Elever, der læser sikkert, kan derimod ofte udvikle deres læsefærdighed betydeligt ved at opøve en højere hastighed.

Som lærer får man endnu flere scorerer at forholde sig til, hvis man også ser på delscorer vedrørende bestemte grupper af opgaver. Men man får også mulighed for at komme mere i dybden med den enkelte elev. I vores ordlæseprøver kan man eksempelvis vælge at se på, hvordan eleven klarer sig for bestemte typer af ord (fra korte og lydrette til lange og/eller uregelmæssige) og dermed få en ide om, hvilke ordtyper der kan gøre en tekst svær for den enkelte elev. Et andet eksempel på delscorer, som vi håber kan uddybe lærerens indsigt, er, at eleverne i prøven *Tekstforståelse Udsagn* for hver læst tekst skal tage stilling til en række udsagn og angive, om de passer med teksten. Blandt disse udsagn ligger de første ret tæt på tekstens ordlyd, mens det ofte kræver inferensdragning (læsning ”mellem linjerne”; se Daugaard, 2015) at tage stilling til de sidste. Dermed kan fejlmønstret give et indtryk af, om både de mest basale og de mere avancerede forståelseskomponenter er på plads.

Fortolkelige resultater: normer og kriterier

Pointen med de ovennævnte scorerer og delscorerer er at gøre prøveresultaterne informative og fortolkelige. Overordnede scorerer er primært informative, hvis de kan sammenlignes med normer, dvs. empirisk baserede oplysninger om, hvordan elever på et givet klassetrin normalt klarer sig i prøven. I samarbejdet med vores forlag har det derfor været af stor betydning, at forlaget har kunnet stå for at indsamle normer fra relativt store elevgrupper fra et bredt udvalg af danske skoler. Ved normbaseret fortolkning får man en pejling på, hvordan elevens eller klassens niveau ligger i forhold til, hvad man typisk ser på klassetrinnet. Det er væsentlig information, men det må understreges, at læsning ikke er en konkurrence – og at cirka halvdelen af en årgang i sagens natur scorerer under gennemsnittet!

Pædagogisk er det efter vores mening vigtigt, at den normbaserede fortolkning suppleres med kriteriebaseret fortolkning.

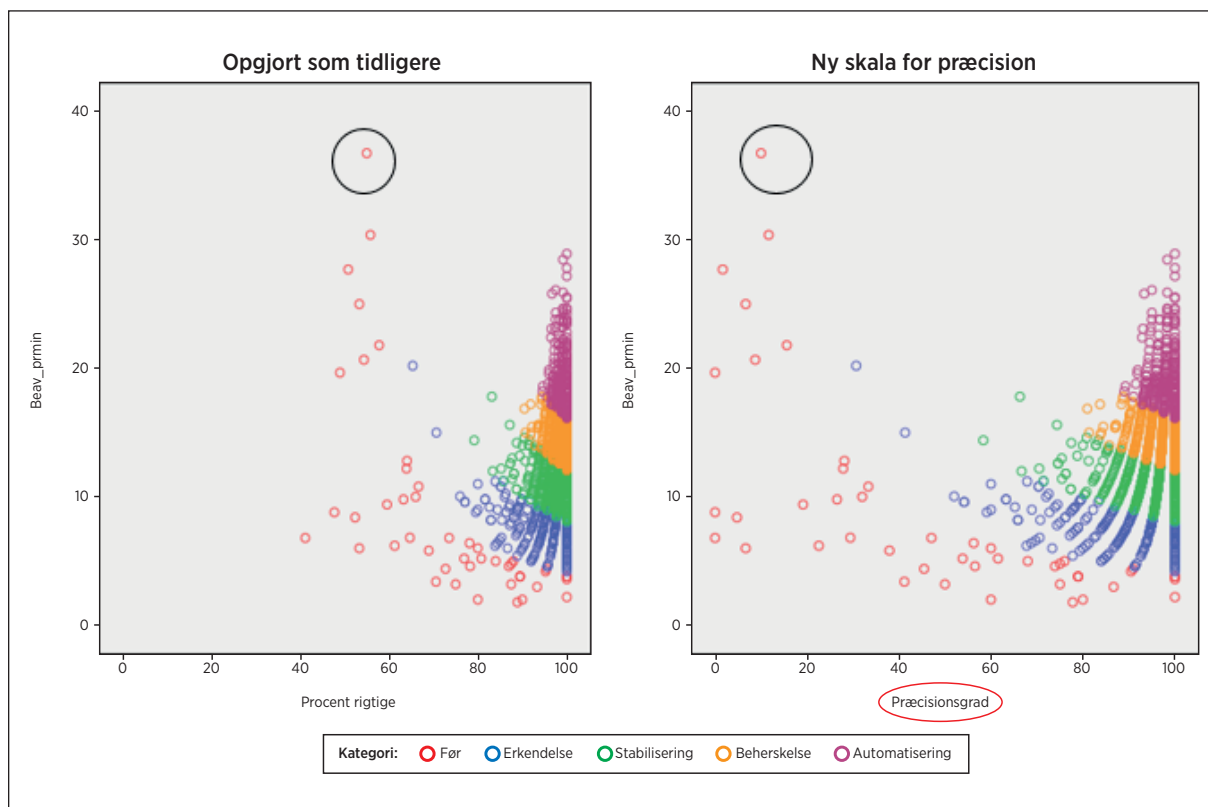
Pædagogisk er det efter vores mening vigtigt, at den normbaserede fortolkning suppleres med kriteriebaseret fortolkning. Det er her delscorerer, som de ovennævnte kommer ind i billedet. Ved kriteriebaseret fortolkning ser man på, hvad eleverne har hhv. let og svært ved – uanset om eleverne er typiske for deres klassetrin eller ej. Man kan eksempelvis spørge, hvad der kendetegner de ord/sætninger/tekster, som eleven læser ubesværet, og hvad der kendetegner de ord/sætninger/tekster, som eleven læser langsommere og/eller mere usikkert. Med den viden bliver det lettere at opstille meningsfulde læringsmål – og lettere at hjælpe eleven med at finde egnede tekster til selvstændig læsning.

Multiple choice-problemet – og vores løsning

Til alle prøverne i serien har vi valgt at bruge opgaver i multiple choice-format. Denne type opgaver har to meget store fordele: Prøver kan afvikles med en hel klasse ad gangen, hvor hver elev arbejder i sit eget tempo. Og opgørelsen af scorerer kan automatiseres, sådan at resultaterne er tilgængelige straks efter prøvens afslutning.

Imidlertid følger der også et klassisk fortolkningsproblem med valget af multiple choice-prøver. Problemet er illustreret i diagrammet til venstre i Figur 1.

Vandret i diagrammet vises procent rigtige i en sætningslæseprøve (med to valgmuligheder pr. opgave), og lodret vises antallet af besvarede opgaver pr. minut. Hver cirkel i diagrammet repræsenterer en elev med en given kombination af præcision og hastighed. Den indcirklede elev har opnået en ganske høj samlet score (faktisk over middel for elevens klassetrin!) ved at besvare opgaverne med høj hastighed – men med en præcision, der kun ligger lidt over 50 %, og som kan være resultatet



Forholdet mellem præcision og hastighed opgjort på to lidt forskellige måder. Data fra normeringen af prøven Sætningslæsehastighed (gengives med tilladelse fra Hogrefe Psykologisk Forlag).

af rene gættesvar. Problemet er med andre ord, at man i multiple choice-prøver kan opnå tilsyneladende fine scorere – uden at man nødvendigvis forstår, hvad man læser!

Dette problem har vi indført en løsning på i de senest udkomne prøver til 6. til 8. klasse (se diagrammet til højre i Figur 1). Her trækker vi nemlig procenten af forkerte svar fra procenten af korrekte svar, sådan at scoren for en elev med 50 % rigtige og 50 % forkerte bliver 0. Denne score kalder vi Præcisionsgraden. Fordelen ved denne justerede præcisionsscore er, at elever med gætteadfærd får scorere tæt på nul. Når Præcisionsgraden kombineres med hastighedsmålet til en samlet score, som vi kalder Læseflow, får elever med gætteadfærd også på dette mål en score tæt på nul. Dermed bliver der overensstemmelse mellem elevernes præstationer og deres scorere, også hvis de gætter sig frem. Interesserede kan finde beregningsformlerne i prøvevejledningen (Møller & Juul, 2017)

Et kategorisystem

Som det fremgår af Figur 1, har vi valgt at inddelle elevernes resultater i resultat kategorier, som er baseret på kriterier for præcision og hastighed. Elever, der har lav præcisionsgrad eller svarer usædvanlig langsomt, placerer sig i, hvad vi kalder Før-fasen (elever markeret med rødt i Figur 1). Resultatet tyder på, at de endnu ikke er så langt i deres udvikling, at de kan klare de krav, prøven stiller. I den anden ende af kategorisystemet har vi elever i kategorien Automatisering – det er elever, som kombinerer en høj præcisionsgrad med en høj hastighed (markeret med lilla i Figur 1). Her tyder resultatet på, at de klarer prøvens krav stort set uden problemer.

Ideen med kategorisystemet er at gøre resultaterne mere overskuelige for læreren. Ved gruppearbejde kan det i mange sammenhænge være hensigtsmæssigt at samle elever, der scorer i samme kategori. Hensigten var, at de betegnelser, vi valgte at give kategorierne, skulle være informative – men vi har erfaret, at det kan give en vis forvirring, hvis det ikke understreges tydeligt, at de er baseret

på kriterier, der er uafhængige af elevens klassetrin. Figur 2 viser som eksempel den procentvise fordeling af kategoriscorer for *Sætningslæseprøve 1*. Umiddelbart kunne man tro, at en score i den midterste kategori, Stabiliserings-fasen (grøn farve i Figur 2), pr. definition må være utilfredsstillende. Imidlertid er et resultat i denne kategori decideret fremragende, hvis der er tale om en elev, der lige er begyndt i 1. klasse. Og i starten af 2. klasse er et resultat i denne kategori helt normalt og ikke specielt alarmerende.

Betegnelsen *Stabilisering* angiver, at der i forhold til prøvens krav er et betydeligt rum for fortsat udvikling, mens *Automatisering* angiver, at med de krav, der stilles i prøven, er eleven ved at være i mål. Elever, der placerer sig i kategorien *Automatisering* i *Sætningslæseprøve 1*, placerer sig typisk i en lavere kategori i den mere krævende *Sætningslæseprøve 2*.

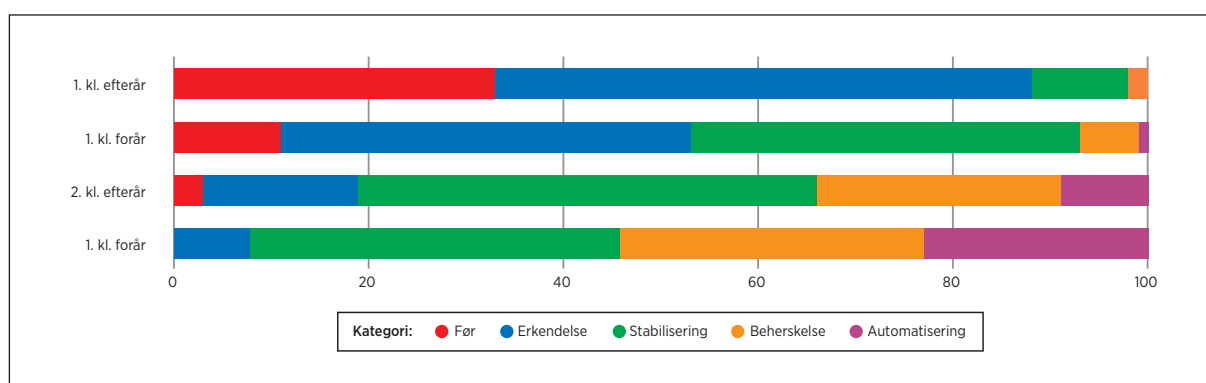
Digitale muligheder

Med de digitale versioner af prøverne blev det nødvendigt at normere prøverne på ny – hvilket vil sige, at de igen blev afprøvet på en stor og nogenlunde repræsentativt udvalgt elevgruppe – fordi afviklingsvilkårene jo var lidt anderledes. Det var på tale at ændre på kategorigrænserne, fordi svartiderne i snit viste sig at være lidt kortere, når eleverne skulle svare med klik og ikke som tidligere med blyant. Med digital afvikling blev det en smule nemmere at få en score i kategorien *Automatisering*. For at undgå forvirring valgte vi dog at bibeholde de allerede indførte kategorigrænser.

Ved papirprøverne var det et vilkår, at eleverne nogle gange overså en opgave (eller et helt opslag i prøvehæftet), hvilket kunne trække det samlede resultat lidt ned. Med digital afvikling undgår vi overspring, for man kan først gå videre, når man har afgivet sit svar. Samtidig er det også lettere for eleverne at holde fokus, fordi der kun er en enkelt opgave på skærmen ad gangen. Det var dog en længere diskussion, om ikke eleverne skulle have lov at vælge ”ved ikke”, når de følte sig usikre på svaret. Nogle elever bryder sig ikke om at svare, hvis de ikke føler sig 100 % sikre i deres sag. Når vi endte med at vælge ved-ikke-muligheden fra, var det for at stille eleverne så lige som muligt. Elever, der ville have svaret ”ved ikke”, hvis de kunne, viser sig ofte at kunne svare lige så godt som elever, der gladeligt giver deres bedste bud. Desuden ville en ved-ikke-knap betyde, at nogle elever ville klikke sig igennem en hel prøve uden overhovedet at afgive svar – og i så fald bliver hele prøveafviklingen jo et spild af både tid og penge.

Med digital afvikling er der mulighed for at gøre endnu flere resultater tilgængelige for lærerne end tidligere. Fx kunne vi i princippet vise svartiden for hver enkelt opgave i ordlæseprøverne. Men her er det en fare, at man som lærer bliver druknet i informationer og i valgmuligheder, som der skal sættes flueben ved. I forlængelse af dette vil vi slutte med at gentage vores indledende pointe:

De valg (og fravalg), man tager som forfatter af en læseprøve, betyder uvægerligt, at man reducerer – og risikerer at overse noget interessant. Men et vigtigere hensyn er det, at både afvikling og



Den procentvise fordeling af kategoriscorer i *Sætningslæseprøve 1*. Data fra 2015-normeringen (gengives med tilladelse fra Hogrefe Psykologisk Forlag). Der er i 2018 udgivet opdaterede normer.

