

Måling af flydende læsning med øjenbevægelser i skolen

Eye-tracking kan under selvstændig læsning i 4. kl. måle variation i læse- og stavefærdigheder.

Af Sigrid Klerke & Stine Fuglsang Engmose

Korrekt citering af denne artikel efter APA-systemet

(American Psychological Association System, 7th Edition):

Klerke, S. & Engmose, S. F. (2024). Måling af flydende læsning med øjenbevægelser i skolen. Eye-tracking kan under selvstændig læsning i 4. kl. måle variation i læse- og stavefærdigheder. *Learning Tech - Tidsskrift for læremidler, didaktik og teknologi*, (14), 98-134. DOI: 10.7146/lt.v9i14.137119

Abstract

Eye-tracking-teknologien er blevet tilgængelig for måling af elevers læsning i skolen. I en skolehverdag skal eye-tracking-data kunne omsættes til pålidelige mål med praktisk relevans, fx til undervisnings-evaluering. Dette studie præsenterer øjenbevægelsesmålet fokusord, der er udviklet til at opmærke nedbrud i elevers flydende læsning og dermed et forsøg på at oversætte eye-tracking-data til læsefaglig indsigt. Studiet er designet til at validere målets pålidelighed som læse-/stave-indikator og endvidere målemetodens robusthed. 84 elever fra fem 4.-klasser deltog, hvoraf data fra 68 elever indgår. Udover elevernes eksisterende resultater i læse-/stavetest, optog vi deres øjenbevægelser under selvstændig højtlesning. Vi finder stærke korrelationer mellem øjenbevægelser og læse-/stavetestresultater, som indikerer målemetodens robusthed. Videre korrelerer fokusord med klassiske øjenbevægelsesmål, hvilket er tegn på samstemmende validitet. Endelig forklarer fokusordene unik variation i elevernes læseforståelses-effektivitet og stavefærdighed ud over, hvad det stærkest korrelerede klassiske mål, fikseringstid, forklarer. Resultaterne understøtter, at fokusordsprocent er robust og tæt forbundet med læse- og stavefærdigheder.

Availability of eye-tracking technology has increased, also for measuring pupils' reading at school. In a school day, eye-tracking data must be able to be translated into measurements that are both reliable and have practical relevance, e.g. for evaluating teaching. This study presents the eye movement measure focus-words designed to mark breakdowns in students' reading fluency and hereby translate eye-tracking data into reading insights. The study is designed to validate the reliability of the measure as an indicator of reading and spelling proficiency and the robustness of the measurement method in practice. A total of 84 pupils from five fourth-grade classes participated, of which data from 68 pupils was included. Together with the students' existing test scores in spelling and reading comprehension, we collected recordings of their eye movements while they read aloud independently. We find strong correlations between eye movements and reading/spelling test results, which indicates that the measurement method is robust. Further, focus-words correlate with classic eye movement measures, which is a sign of co-validity. Finally, focus-words explain unique variation in students' reading comprehension and spelling proficiency beyond what the most strongly correlated classical measure, fixation time, explains. The results support that focus-word-percentage is a robust metric closely related to literacy.

Måling af flydende læsning med øjenbevægelser i skolen

Eye-tracking kan under selvstændig læsning i 4. kl. måle variation i læse- og stavfærdigheder.

Indledning

Det er veletableret, at der er en sammenhæng mellem læsefærdigheder og muligheder for at deltage i samfundet. Den internationale læseundersøgelse PIAAC (Rosdahl, Fridberg, Jakobsen & Jørgensen, 2013) peger på, at gruppen af unge og voksne danskere, som ikke er funktionelle læsere, har større sandsynlighed for at nå kortere i uddannelsessystemet, for at stå uden for arbejdsmarkedet, for ikke at tage en efteruddannelse og højere risiko for ringe helbred. Disse resultater bliver bekræftet i den nationale undersøgelse af voksnes basale færdigheder (Larsen, Jakobsen & Rosdahl, 2022). Også en undersøgelse, der følger op på danske unges uddannelsesstatus fire år efter, de som 16-årige har deltaget i PISA-undersøgelsen, finder en stærk sammenhæng mellem læsefærdighed og uddannelsesstatus (Andersen, 2005). Andersen (2005) finder, at unge, som ikke er i uddannelse og heller ikke har gennemført en ungdomsuddannelse i løbet af de fire år, har lavere læsefærdighed. Samtidig er gruppen af unge, som ikke er funktionelle læsere, uændret fra PISA-undersøgelsen i 2009 og frem til 2018. Andelen ligger stabilt omkring 15 % (Christensen, 2019).

Det er i grundskolen, eleverne lærer at læse. Det afspejler sig i kompetence-, færdigheds- og vidensmålene for danskfaget (Børne- og undervisningsministeriet, 2019). Efter 2.kl. er kompetencemålet for læsning, at eleverne kan læse enkle tekster sikkert og for færdigheds- og vidensområdet afkodning er målet, at eleverne ved slutningen af 2.kl. kan læse ord i tekster til klassetrinnet sikkert. Der er altså fra grundskolens start et fokus på at udvikle sikker afkodningsfærdighed, som kan anvendes til at læse tekster. Senere i grundskolen udvides fo-

Sigrid Klerke, Det Kongelige Akademi,
& Stine Fuglsang Engmose, Professionshøjskolen Absalon

kus i målene for færdigheds- og vidensområdet afkodning mod, at eleverne også udvikler en hurtig og automatiseret afkodningsfærdighed med umiddelbar ordgenkendelse. I sine bøger “Læsning og læseundervisning” (2016) samt “Læsevanskeligheder” (2021), gennemgår professor Carsten Elbro årtiers international forskning i læseudvikling og -undervisning. Elbro (2016 og 2021) peger på, at direkte og systematisk undervisning i skriftens lydprincip er afgørende for udviklingen af sikker afkodnings- og stavefærdighed, herunder hører omfattende øvelse med at læse tekster af passende sværhedsgrad. Dette gælder i særdeleshed for elever med ordblindhed, som har svært ved at tilegne sig sikker og automatiseret afkodnings- og stavefærdighed (Elbro, 2021). Læsning er dog mere end afkodningsfærdighed. Læsning rummer også det at forstå de tekster, man læser. For at kunne tilpasse læseundervisningen til den enkelte elev og identificere elever, som ikke udvikler deres læsefærdighed, og som måske har brug for mere intensiv læseundervisning, har folkeskolen haft Folkeskolens Nationale Test i Dansk for 2., 4., 6., og 8.kl. Testene har været kritiseret for at give ikke-gyldige informationer om elevernes færdigheder (Ravn, 2015, Kousholt, 2015), og de er ikke længere i brug, men nye nationale test i dansk er på vej (Børne- og undervisningsministeriet, 2022). Ved siden af de obligatoriske test anvender mange skoler supplerende standardiserede pædagogiske og diagnostiske test. Disse kan fx følge elevernes læsefærdighed over tid, pege på særlige nuancer i fx afkodningen, stavningen eller læseforståelsen den enkelte elev er usikker i, og sammenligne den enkelte elevs færdigheder med en normgruppe. Standardiserede test er på den måde et væsentligt redskab i den løbende tilpasning af læseundervisningen til den enkelte elev.

Der er dog også ulemper ved traditionelle, standardiserede test. Man kan kritisere formatet i disse test for at være “langt” fra det arbejde, eleverne normalt laver, når de læser tekster i skolens fag. For at kunne sammenligne elevernes færdigheder med en normgruppe, er de standardiserede test designet med en fastlagt ramme, for hvornår og hvordan de kan gennemføres. Testsituationen er på den måde ikke designet til at ligne almindelig undervisning. Test af denne type er fx født med den begrænsning, at man ved gentestning inden for kort tid ofte vil se, at eleverne scorer højere end første gang, fordi de kender testen¹. En anden bekymring er, at testsituationen kan være ubeha-

1 Standardiserede test kommer dog ofte med information om test-retest reliabilitet, så man har en fornemmelse af denne øvelseseffekt.

gelig for nogle elever. Dette kan fx ses i en national spørgeskemaundersøgelse med besvarelser fra 768 lærere om brugen af Folkeskolens Nationale Test (Bundsgaard og Puck 2016). Svarene fra lærerne i denne undersøgelse tyder på, at mange lærere oplever, at der er elever, der bliver kede af det under testen, og at testen ikke opleves meningsfuld for de svageste elever. Endvidere er denne type af test ofte opbygget, så de kan indfange variationen i en klasse fx mellem den svageste og stærkeste læsere. Det betyder, at en stor del af eleverne som gennemfører testen vil opleve, at de skal prøve at løse opgaver, som er for svære for dem, eller at de kun når halvdelen af testens opgaver. Det er ikke sikkert, at de elever, som scorer lavt på testen, oplever arbejdet med testens opgaver som et nederlag, men hvis de gør, så kan netop gentagne erfaringer med nederlag påvirke elevernes selvbillede i faget negativt. I en gennemgang af litteraturen om selvbillede og elever med læsevanskeligheder konkluderer Lena Swalander (2012), at der er en vedvarende sammenhæng mellem problemer med læsning og selvbilledet og en øget risiko for, at denne gruppe af elever får negative følelser i forbindelse med præstationer (Swalander, 2012). I en sammenligning af 78 skoleelever med ordblindhed og 77 jævnaldrende uden, finder Alexander-Passe (2008) signifikante forskelle mellem de to grupper, hvor eleverne med ordblindhed oplever sig mere pressede. Det kommer særligt til udtryk i interaktionen med lærere, prøvesituationer og færdighedstest. Eleverne fortæller om angst, generthed, ensomhed, kvalme, rystende hænder og bankende hjerte. Flere studier finder lignende sammenhænge mellem studerende med ordblindhed og angst ved prøvesituationer (Caroll & Iles, 2006, Nelson et al., 2015). Disse undersøgelser peger på, at der for gruppen af elever i læsevanskeligheder er grund til at tage risikoen for at skabe negative erfaringer med læsning i betragtning, når de skal gennemføre test, der stiller høje krav til læsefærdigheden.

I forskning i læsning og læseudvikling er øjenbevægelser veletablerede indikatorer for kognitiv processering (Rayner, 2012, Holmqvist, et al., 2011). I takt med at eye-tracking er modnet som teknologi, og der er udviklet pålidelige ikke-begrænsende og brugervenlige instrumenter til at optage øjenbevægelser, er interessen også steget for at anvende øjenbevægelser i læseevaluering (Gran Ekstrand et al., 2021, Benfatto et al., 2016, Klerke et al., 2018, Bingel et al. 2018). Forventningen er, at eye-tracking-teknologien kan give nye muligheder i evalueringen af læsning. Særligt kan metoden imødegå nogle af ulemperne ved de standardiserede test, fordi øjenbevægelser kan optages under uformel læsning af egne tekster. Formodningen er, at sådanne mindre formelle optagelser kan bruges til at evaluere læsefærdighed løbende, så læreren ad den vej kan få indsigter, som kan danne grundlag for den løbende tilpasning af læseundervisningen til den enkelte elev.

Traditionelt er eye-tracking-mål i læseforskning blandt andet blevet brugt til at udpege aspekter af skriftsproget som systematisk kan påvirke læseres øjenbevægelser, fx ved at føre til længere eller flere fikseringer på de ord som kræver ekstra processering. Derfor forventer vi i dag at se effekter af ord der er sjældne, lange eller har en overraskende grammatisk rolle. Effekterne ses konsistent på de fikseringsbaserede mål som gennemsnitlig fikseringstid og varigheden af den første fiksering per ord (Rayner, 2012). Et andet klassisk fokus har været at lede efter øjenbevægelsesmål som pålideligt skelner aspekter ved læsere som fx ordblindhed (Benfatto et al., 2016, Franzen et al., 2021). Eksempler på mål der kan bidrage til at skelne forskelle i læsestrategier og -færdigheder, er andelen af fikseringer brugt på at genlæse teksten, andelen af oversprungne ord og hvilke retninger øjnene typisk flytter sig i. Mål, der aggregerer elevens sammenhængende læsning, er især anvendt til dette formål, fordi læseprocesser relateret især til forståelsen fører til en wrap-up-effekt, hvor øjnene ofte dvæler længere ved ord, der står ved sætningsgrænser (Meister, et al., 2022, Rayner, 2012, Franzen et al., 2021). Selvom både klassiske og moderne mål er blevet brugt til at skelne læseres færdigheder og teksters sværhedsgrad, er forskelle i øjenbevægelser en effekt af og ikke en årsag til forskelle i læsning. Det betyder også at det ikke er didaktisk meningsfuldt hverken at forhindre genlæsning eller at forsøge bevidst at ændre på fikseringstider for at styrke selvstændig læsning (Holmqvist et al., 2011, Rayner, 2012)². Ligeledes er målene ikke designet til at tolke noget om det enkelte individ, hvilket har den konsekvens, at det ikke er oplagt, hvordan praktikere kan omsætte varigheden af første fiksering per ord til undervisningstiltag.

Der mangler altså fortsat en praktisk og metodisk modenhed; dels skal det afklares, i hvilket omfang øjenbevægelser, også afspejler læseres læsefærdigheder under hverdagslæsning på egen hånd, og dels er der brug for mål, der kan tolkes uden specialopklæring i brug af eye-tracking. Ideelt set skal målene kunne tolkes og anvendes med den specialviden, læsevejlederen allerede har.

Dette studie tager hul på at adressere begge udfordringer. Først korrelerer vi øjenbevægelsesmål fra forskningslitteraturen målt i en realistisk skolekontekst med standardiserede testresultater for

2 Holmqvist et al., 2011 indeholder en grundig gennemgang af et meget bredt udvalg af klassiske eye-tracking mål, deres ophav, varianter og hvordan de typisk anvendes og tolkes. Rayner, 2012 opsummerer specifikt læseforskning med eye-tracking.

læseforståelse og stavning, som er, set ud fra forlagenes testudbud, standardiserede test, der ofte anvendes, som redskaber til klasse-screening på mellemtrinnet. Videre præsenterer vi en kvantificeret udgave af målet *fokusord*, der automatisk opmærker nedbrud i flydende læsning. Opmærkingen er udviklet, så de udpegede ord tillader læsevejlederen at fortolke disse fokusord i analogi til fejllæsninger i konventionel læseevaluering som fx. *running record* (Wangebo, 2011, D'Agostino et al., 2021). Ligesom fejllæsninger, skaber fokusord et udgangspunkt for evalueringen af en elevs læsning, men i stedet for manuelt opmærkede fejl, udpeger fokusord automatisk de *ord, der bremser elevens flydende oplæsning* målt med øjenbevægelser. Det er didaktisk meningsfuldt at kunne udpege og afspille de konkrete ord, der kostede eleven ekstra opmærksomhed, for at planlægge og evaluere læseundervisningen. Her undersøger vi om procenten af fokusord, i lighed med fejlprocenten (cf. Nielsen et al., 1992), afspejler læse- og stavefærdigheder.

Vi har operationaliseret projektet i følgende tre forskningsspørgsmål og hypoteser:

1. Hvor robust er den kendte forbindelse mellem læse- og stavefærdigheder og eye-tracking-mål, når elever læser sammenhængende tekster med eye-tracker i et skolevenligt-setup?



Vores overordnede hypotese er, at de klassiske eye-tracking-mål ikke vil være lige robuste i dette setup, hvor der må forventes mere støj sammenlignet med kontrollerede laboratorie-designs, hvor målene ellers normalt korrelerer med læsefærdigheder.

2. Korrelerer eye-tracking-målet fokusord med læse- og staveresultater i samme grad som de klassiske eye-tracking-mål?



Hypotesen er, at det nye eye-tracking-mål, fokusord, afspejler mængden af ord, som eleven har svært ved at afkode, og derigennem afspejler både læse- og stavefærdigheder. Der er flere måder at kvantificere målet fokusord. Vi bruger korrelationsanalysen til at udpege den variant som er numerisk stærkest korreleret med læse-/stavemålene som den mest robuste. Vi vælger altså at afgøre dette empirisk da vi ikke har tilstrækkeligt teoretisk belæg for at favorisere en enkelt variant på forhånd. Dette betyder at vi har ekstra mange sammenligninger som gør det sværere at opnå pålideligt signifikante resultater.

3. Forklarer fokusord unik variation i elevernes læseforståelse og stavefærdighed ud over den variation, der forklares af det traditionelle eye-tracking-mål, der er numerisk stærkest forbundet med læse- og stavemålene?

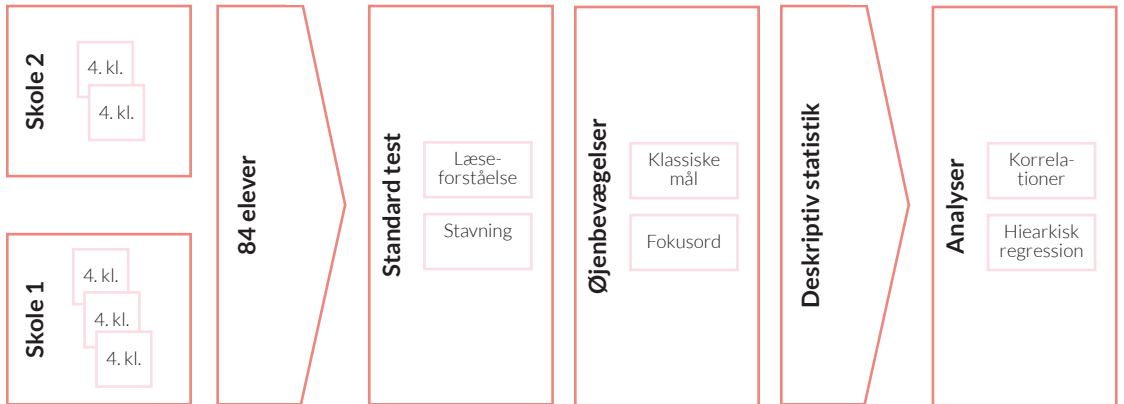


Her er hypotesen, at fokusordenes særlige følsomhed for afbrud i flydende læsning kan afspejle unikke aspekter af læseforståelse og stavefærdighed, som ikke allerede afspejles i det klassiske eye-tracking-mål, der er numerisk stærkest korreleret med læse- og stavefærdigheder. Hvis det er tilfældet understøtter dette studie, at fokusord er mere og andet end det klassiske eye-trackingmål, og samtidig er en pålidelig læse-/stave-indikator.

Metode

Undersøgelsens design er illustreret i Figur 1. Vi gennemfører korrelationsanalyser og hierarkiske regressionsanalyser med data fra to typer af måleredskaber: Dels standardiserede test af læseforståelse og stavning, og dels optagelser af øjenbevægelser under tekstlæsning (se figur 1).

Figur 1.
Undersøgelsens design



Deltagere

To skoler fra to kommuner deltog med i alt 84 elever fra fem forskellige klasser på fjerde årgang. Vi har inkluderet alle, der deltog både i mindst en af de to standardiserede tests og i læsning med eye-tracker. Læsninger fra tre elever, der har brugt oplæsningsstøtte under læse- eller stavetesten blev ekskluderet. Fire læsninger er ekskluderet, hvor elever har angivet, at de kendte bogen fra tidligere. På baggrund af tre foruddefinerede kriterier for datakvalitet (se bilag 1) blev ni læsninger ekskluderet. I fire af tilfældene var eye-trackingen for upræcis, i fire tilfælde var der udfald i data fra eye-tracking, undervejs og i et tilfælde havde en elev stillelæst. Tabel 1 viser antallet af elever før og efter eksklusion.

Tabel 1.

Antal deltagere før og efter eksklusion (se tekst).

	Læsetest	Stavetest	Eye-tracking
Antal elever deltaget	84	81	84
Antal elever bibeholdt	78	77	68

Gennemførelse

For at minimere negative oplevelser, har vi tilstræbt at indsamle så lidt testdata som muligt. Derfor er de standardiserede test i undersøgelsen netop de test af læseforståelse og stavning, som skolerne alligevel gennemførte, så disse blev gennemført som en del af kommunernes generelle læseindsats 1-3 måneder før læsning med eye-tracker blev gennemført. Øjnbevægelsesdata er indsamlet i maj måned 2021.

Under måling af øjenbevægelser læste eleverne tre tekster højt ved en computer med eye-tracker. Under indsamlingen af øjenbevægelsesdata har vi søgt at minimere elevernes tid væk fra klassen ved at planlægge deres deltagelse sammen med skolens læsevejleder. Eleverne blev derfor i klassen introduceret til at læse e-bøger med eye-tracker via en mundtlig præsentation og en video der viste login, kalibrering og læsning med eye-tracker ([intro-video](#)). Eleverne fik ved gennemførelsen af deres læsning hjælp ved tekniske vanskeligheder, men læste ellers højt for sig selv på egen hånd. Hver læsning sluttede efter et fast tidsinterval, hvilket undlod at udstille nogen som hurtigere eller langsommere læsere. Seks bøger var udvalgt til forsøget. Alle elever fik læst de seks bogtitler op, og blev bedt om at udfylde et ark med et billede af hver bog, hvor de skulle sætte kryds ud for de bøger, de tidligere havde læst i. Læsningerne blev gennemført over en dobbeltlektion, hvor 2-4 elever ad gangen læste samtidig i et separat lokale under opsyn. Hver elev gennemførte tre læsninger, en demo-læsning på 1 minut (lix 6), en læsning i en forudbestemt bog i 3 minutter (lix 18) og en læsning i en selvvalgt bog i 2 minutter (lix 11-25). Vi analyserer den

forudbestemte læsning på 3 minutter. Det gør vi for at give sammenligningen med de standardiserede test bedst mulige vilkår i denne første undersøgelse af fokusord. Bogen til den forudbestemte læsning var derfor valgt, så teksten matchede den standardiserede læseforståelsestest på ord- og sætningslængde. Dermed er data i denne undersøgelse fra den del af læsesituationen, hvor alle elever læser den samme tekst. Tidtagningen med eye-tracker følger øjenbevægelserne og stopper midlertidigt, hver gang eleven ser væk fra teksten i mere end et halvt sekund. Når den planlagte læsetid er gået, bliver det umuligt at bladere videre i bogen, men eleven vælger selv hvornår bogen lukkes. Det betyder, at hver optagelse varer lidt længere end hhv. 1, 3 og 2 minutter, hvis eleven enten har kigget væk fra teksten, har brugt ekstra tid på at læse sidste side færdig, eller hvis eye-trackerens kalibrering er meget upræcis. Første side i hver e-bog er et billede af forsiden af bogens originale omslag, og første side regnes derfor altid kun med som et billede, der højst kan tælle for et halvt sekund af den samlede læsetid. Eleverne beskrev overvejende opgaven som let.

Det er denne læsesituation vi omtaler, som skolevenligt-setup. Dette skal ses i lyset af det klassiske eksperimentelle setup, hvor læserens hoved fx støttes til at være i samme position under hele læsningen eller næsten identiske stimuli gentages mange gange.

Forud for deltagelse blev forældrene informeret i et kort skriv, og der blev indhentet informeret samtykke fra forældre/værge til deltagelse.

Anvendte skriftsproglige mål

Tekstlæseprøve 5 (Møller, 2013) er en læseforståelsesprøve med fokus på elevens evne til at drage inferenser. Teksten er bygget op af almindelige ord for at mindske ordforrådets betydning. Testen bruger to spørgsmålsformater 1) cloze i teksten, hvor eleven skal vælge det rigtige ord og 2) ja/nej spørgsmål for hver 3. til 4. side. Testen har 44 items. Det er muligt at sammenligne klassen og eleverne med normer og kriteriebaserede kategorier (se bilag 1 for flere detaljer).

Vi bruger tre mål fra læseforståelsesprøven, da vi ønsker at belyse sammenhængen mellem fokusord og mål af læseforståelse der er mere eller mindre følsomme for afkodningsfærdighed. Det første mål er besvarelsestiden, som er det tætteste vi kommer et rent mål af læsehastighed i testen. Det andet mål er antal rigtigt bevarede forståelsesspørgsmål per minut. Vi bruger dette som effektivitetsmål, hvor scoren både påvirkes af, hvor sikker forståelsen er, og også hvor hurtig afkodning og forståelse foregår. Det sidste mål er testens præcisionsmål, der angiver procent rigtigt besvarede forståelsesspørgsmål. Dette mål afspejler sikkerheden i elevens forståelse. Denne sikkerhed kan dog også være påvirket af usikker afkodning, fx ved at langsom afkod-

ning gør det svært for eleven at huske information fra teksten, eller fordi upræcis afkodning giver eleven et usikkert grundlag at forstå teksten på.

Staveprøve 3 (Juul, 2019) er en indsætningsdiktat med 36 items. De er udvalgt til at være almindelige ord med almindelige grammatiske suffikser. Det er muligt at sammenligne klassen og eleverne med normer og kriteriebaserede kategorier (se bilag 1 for flere detaljer). Vi bruger målet andel korrekt stavede ord, som teoretisk set er mest sammenligneligt med elevens afkodningspræcision. Ordafkodning og stavning er stærkt korreleret. Fx finder Caravolas og kollegaer i et longitudinelt studie, hvor de følger engelsksprogede elever fra midten af børnehaveklassen til slutningen af 2.kl., at korrelationen allerede fra slutningen af børnehaveklassen er stærke og at den ved slutningen af 2.kl. er meget stærk med en Pearsons korrelation på 0,81 (Caravolas m.fl., 2001). Færdighederne er altså tæt forbundne, men ikke den samme færdighed.

I begge test rapporterer forfatterne om høj intern konsistens for skalaen. Der er ligeledes foretaget Rasch-analyse, hvor items uden Rasch homogeneity er sorteret fra. Det betyder, at vi for begge test kan regne med, at vi kan sortere personerne på baggrund af deres faktiske dygtighed inden for testens område.

Anvendte øjenbevægelsesmål

Det grundlæggende mål af øjenbevægelser under læsning er en fiksering. Kun et lille område midt i synsfeltet står skarpt nok til at se bogstaver og ord tydeligt. For at afkode en hel tekst, må det centrale synsfelt derfor flyttes rykvis over teksten i et tempo og mønster, der tillader læserens afkodning og sprogforståelse at følge med. Fikseringer under oplæsning varer typisk 100-500 ms, og den nedre grænse for at afkodning af bogstaver kan nå at ske, er ca. 50 ms. Et ryk er en saccade og baglæns saccader kaldes regressioner. Regressioner udgør 10-15% af trænedes læseres læsning (Rayner, 1998).

Varigheden af den første fiksering på et ord bliver kodet ind i øjenmuskulernes bevægelsesprogram, inden ordet er set. Den første fiksering afspejler derfor den tid, læseren *forventer* at bruge på at genkende det fikserede ord. Når processeringen kræver længere tid end forventet, opsøger læserens øjne opklarende information automatisk og ubevidst. Da signaler til øjenmuskulaturen tager tid at udføre, kan en planlagt saccade nå at blive udført, selvom et ord viste sig at kræve mere opmærksomhed end forventet. Det giver et mønster, hvor udfordrende ord ofte genbesøges indenfor kort tid, for at tillade afkodningen og sprogforståelsen at indhente øjnene (von der Malsburg & Vasishth, 2011). Fokusord er designet til at udpege ord, der kostede

læseren så meget mere opmærksomhed end ventet, at det kan have afbrudt den flydende læsning. Dels fordi der kan være et mønster i de ord læseren snubler over, og dels fordi læserens brug af afkodnings- og forståelsesstrategier kan vise sig netop her.

- Vi beregner følgende øjenbevægelsesmål: Læsehastighed
- To klassiske mål defineret af fikseringstider
- To klassiske mål der forudsætter sammenhængende tekstlæsning
- Seks varianter af målet fokusord der detekterer sekvenser af fikseringer, som bryder læserens flydende progression.

Læsehastighed er et udbredt læsemål, og eye-tracking måler læsehastigheden med høj præcision, men målet kræver ikke eye-tracking. Læsehastighed er almindeligt brugt til at evaluere elevers læsefærdighed (Elbro, 2021). Hastigheden måles i ord pr. minut, og vi beregner den ud fra antallet af viste ord, fra første tekstside vises og frem til bogen lukkes.

Det er en veletableret indikator for læsning, og indgår ofte som en delkomponent i standardiserede test af generelle læsefærdigheder. I denne undersøgelse fungerer læsehastigheden under eye-tracking som en pålidelighedslæseindikator.: Hvis målets sammenhæng med læse- og stavemålene forsvinder i vores skolevenlige naturlige skole-setup, så er det et tegn på, at der er sket noget andet end læsning under eye-trackingen, og datasættet vil ikke kunne belyse forskningsspørgsmålene. Vi forventer derfor, at målet læsehastighed korrelerer stærkt med både læse- og stavemålene, hvis det er tilfældet.

Gennemsnitlig varighed af ord-fikseringer er varigheden af hver fiksering i gennemsnit. I forsøg med matchede kontroller er der fundet signifikant forlængede ord-fikseringer for både ordblinde voksne (Franzen et al., 2021) og for begynderlæsere (Benfatto et al., 2016). Gennemsnittet kan forvrænges af meget lange fikseringer, hvilket dog modvirkes, når læsetiden øges.

Gennemsnitlig varighed af første fiksering pr. ord (first fixation duration) er gennemsnitstiden af første fiksering på hvert ord. Målet afspejler særligt afkodningsprocesser og læserens forventning til tekstens sværhedsgrad (Holmqvist et al. 2011). Målet har blandt andet været anvendt til at undersøge hvor tidligt ord yderst i det centrale synsfelt kan genkendes – en evne der først udvikles efter tilegnelsen af basale læsefærdigheder (Sperlich et al., 2015). Når læsere udvikler bedre perifer afkodning, giver det højere læseeffektivitet, som bl.a. giver kortere første fiksering pr. ord.

Begge ovenstående mål er klassiske øjenbevægelsesmål, der

ændrer sig med generelle kognitive færdigheder, specifikke læsefærdigheder og aspekter af teksten. Det gælder både under læsning af enkeltord og tekstlæsning (Rayner, 1998, Holmqvist et al. 2011). Vi forventer på den baggrund en signifikant sammenhæng mellem disse mål og med mål som afspejler afkodningsfærdighed. Det er en praktisk svaghed ved begge mål, at bidraget fra fx læsefærdigheder eller ordenes sværhedsgrad er et svagt signal som let drukner i støj fra andre variationer i hhv. læserne og ordene, hvis de ikke kontrolleres på gruppeniveau i et studie. Dette gør målene svære at bruge uden for laboratoriet til evaluering og undervisning af elever.

First pass dwell time og *andel overskimmede ord* (ratio of words skipped during first pass reading) er begge mål, der kan bruges til at måle læsning af sammenhængende tekster. De er følsomme for ordblindhed hos voksne funktionelle læsere (Franzen et al., 2021). Da ordblinde læsers øjenbevægelser ofte er sammenlignelige med yngre læsematchedede kontroller (Benfatto et al., 2016), *forventer vi*, at målene også er følsomme for væsentlige forskelle i læse- og stavfærdighed i vores elevgruppe. *First pass dwell time* er den samlede tid i sekunder, som læseren brugte på at læse fremad i tekstens læseretning. Målet inkluderer al den tid, læseren har set på det senest læste ord, inden et nyt ord blev fikseret, men ikke nogen genlæsning. Andelen af overskimmede ord er andelen af tekstens ord, som slet ikke blev fikseret i løbet af *first pass dwell time*. Begge mål kan afspejle variation i læsestrategier. Tidsmålet, *first pass dwell time*, vil ligge højt for en læser, der bruger tid på bogstavafkodning, og falde, i takt med at afkodningen automatiseres, og med at læserens strategi baserer sig mere på forudsigelse af teksten, fx med perifer ordgenkendelse. Andel overskimmede ord vil omvendt stige, hvis læseren er god til at forudsige teksten og udnytte mere perifer information. Målet ligger lavere for voksne ordblinde end for en kontrolgruppe (Franzen et al., 2021). Begge mål har den svaghed, at de ikke kan skelne almindelige sikre læsere fra usikre læsere, der bruger en strategi, som ofte søger forud i teksten efter ekstra information.

Fokusord er ord, der blev besøgt i en række af fikseringer med en sammenlagt fikseringstid på mindst et sekund. Der kan være fikseringer uden for ordet undervejs, men hver afbrydelse kan højst vare et halvt sekund. Hvis et ord bliver fikseret flere gange og med længere afbrydelser, registreres det som fokusord, hver gang en sekvens af fikseringer opfylder kriteriet for fokusord. Ved at tillade korte afbrydelser mellem fikseringer på det samme ord, bliver målet robust overfor forskellige afkodningsstrategier, som fx. informationssøgning udenfor ordet, hovedbevægelser og grimasser med øjnene samt måleusikkerhed, som fx ved blink. Et helt sekund brugt på et enkelt ord indikerer, at læseren har brugt ekstra opmærksomhed på ordet,

og denne opbremsning i øjenbevægelserne kan være et tegn på, at den flydende læsning er blevet afbrudt, fx af en afkodnings- eller forståelsesudfordring. Et målt fokusord svarer dermed til en sammenhængende sekvens i oplæsningen, der er meningsfuld at afspille og lytte til for nærmere evaluering.

Vi skelner mellem andelen af *læsetiden*, der bidrager til fokusord og andelen af *læste ord*, der blev fokusord. Andelen af læsetid vægter varigheden af afbrydelser højt, mens andelen af ord, der blev fokusord, vægter hyppigheden af afbrydelser højt.

I udviklingsdata fra tidligere optagede læsninger under lignende forhold har vi observeret, at fokusord har en overrepræsentation af ord nær sætningsgrænser. Her kan opbremsningen også skyldes wrap-up-effekten, der er forbundet med forståelsesorienterede stadier i form af syntaktisk og semantisk processering af hele sætningen (Meister, et al., 2022). Her er det derfor mindre sandsynligt at opbremsningen skyldes afkodning og forståelse af det konkrete fokuserede ord. For at isolere de opbremsninger der primært skyldes ordprocessering fra de opbremsninger der også kunne skyldes sætningsprocessering, inddeler vi de identificerede fokusord i hhv. *sætningsmediale fokusord* og *fokusord ved sætningsgrænser*. I alt sammenligner vi seks varianter af fokusordsmålet, for at afgøre empirisk hvilken der bedst afspejler variation i elevernes læse og stave-færdigheder.

Da der er risiko for, at læsninger med nul fokusord dækker over særligt støjfyldt data, ekskluderes de fra at indgå i denne analyse (se bilag 1).

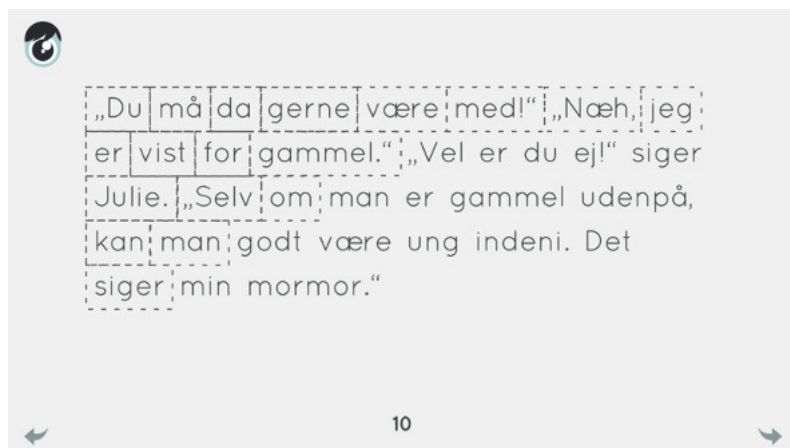
Optagelse af øjenbevægelser

Elevernes øjenbevægelser blev optaget med en Tobii 4C tracker ved en samplingfrekvens på 90Hz på skolernes egne bærbare computere. Teksten blev præsenteret i fonten Quicksand med op til 7 linjer ad gangen, og den blev vist i et centralt område på maksimalt 80% af skærmens bredde og 60% af højden. Den store margin undgår tracking i yderpositioner hvor præcisionen er lavest. Når der er billeder i teksten, deles det centrale areal mellem billedet til højre og teksten til venstre. I hjørnerne af skærmen vises nederst knapper til at bladre i bogen. Knapperne aktiveres med museklik, piletaster eller uafbrudt øjenfiksering i 2 sekunder. I skærmens øverste hjørner findes til venstre en menu-knap i form af et logo. En timer, der viser rest-læsetiden, vises, når læseren kigger eller peger i øverste højre hjørne.

Fikseringer regnes for at lande på et ord, hvis de lander i et rektangulært areal der dækker ordet plus en ikke-overlappende margin på hhv. 1/2 linjeafstand vertikalt og 1/2 ordmellemlinje horisontalt (se Figur 2). Det betyder at alle fikseringer i det centrale tekstområde altid knyttes til et af de viste ord.

Figur 2.

En side i demo-læsningen.



De stiplede kasser rundt om ordene viser, hvordan de optagede fikseringer fordeles mellem tekstens ord. Stiplingerne vises ikke under læsningen.

Analysemetoder

I korrelationsanalyserne bruger vi Spearmans rangkorrelation til at sammenligne målene af læseforståelse og stavefærdighed med klassiske og nye eye-tracking-mål. Det tillader os at detektere, om der er monoton sammenhæng mellem to mål, også hvis et mål nærmer sig et plateau. Det er fx tilfældet, hvis der er et stort fald i læsehastighed mellem elever med hhv. to og fire stavefejl, men kun et lille fald i læsehastighed fra 10 til 12 stavefejl. Vi undersøger altså, om en ændring i et øjenbevægelsesmål afspejler en ændring i et læse- og stavemål, uanset om ændringer i fx. læsehastighed kan oversættes til en proportional ændring i stavefejl.

En tommelfingerregel for styrken af sammenhænge er, at Pearsons korrelationer over 0,5 og under -0,5 er moderate, over 0,7 og under -0,7 er stærke, samt over 0,9 og under -0,9 er meget stærke (Hinkle, Wiersma og Jurs, 2003). To variable med en Pearsons korrelationskoefficient på over 0,7 eller under -0,7 deler ca. 50 % af variansen med hinanden. Når to mål deler meget varians, er det et tegn på, at de i høj grad måler noget ens. Derfor kan stærke sammenhænge tolkes som et tegn på samstemmende validitet mellem to mål.

For Spearmans rho findes der ikke samme retningslinjer for sammenhængen mellem koefficientens styrke og specifikke værdier af koefficienten. Værdien har dog tendens til at være meget sammenlignelig med, og lidt lavere end Pearsons r (Laerd Statistics, 2018). Som for

Pearsons korrelationskoefficient veksler værdien af Spearmans rho mellem -1 og 1. Jo tættere koefficienten er på 1 eller -1, desto stærkere er sammenhængen mellem rangordningen i de to mål, og jo tættere på 0, desto svagere er denne sammenhæng (Laerd Statistics, 2018). Så selvom der ikke findes faste regler for, hvor høj Spearmans rho skal være, for at en sammenhæng er meget stærk, så minder værdierne om Pearsons korrelationskoefficient. Derfor vælger vi at trække grænsen for stærke sammenhænge ved Spearmans rho værdier over 0,6.

Når vi laver signifikanstest af mange korrelationer, øges risikoen for at finde tilfældigt signifikante forskelle. Dette tager vi højde for ved dels at Bonferroni-justere signifikansniveauet og dels rapportere 95%-konfidensintervaller for korrelationerne. Intervallets bredde giver et indtryk af hvor præcist den observerede korrelation kan forventes at afspejle populationen og grænseværdiens afstand fra 0 giver et indtryk af korrelationens pålidelighed. Hvis 0 er indenfor intervallet, er korrelationen ikke pålidelig.

I hierarkiske regressionsanalyser undersøger vi om det nye eye-tracking-mål, som er designet til at kunne anvendes i praksis med et evaluerings- og didaktisk sigte, forklarer andet og mere af målene af læseforståelse og stavning end de klassiske eye-tracking-mål, hvilket er et udtryk for, at det nye og de klassiske eye-tracking-mål ikke er identiske mål. I de hierarkiske regressionsanalyser kan vi introducere klassiske mål i analysen før det nye eye-tracking-mål. Hvis det nye mål, på dette trin, bidrager signifikant til at forklare variation i målene af læseforståelse og stavning, så understøtter det, at det nye mål afspejler variation i elevernes læse- og stavefærdighed, som ikke afspejles af klassiske eye-tracking-mål. Det understøtter, at selvom det nye mål er stærkt forbundet til klassiske mål, er de ikke identiske, og at det nye mål samtidig bidrager til at forklare mere variation i målene af læseforståelse³ og stavning end det klassiske eye-tracking-mål.

Vi anvender det ene af de to overordnede typer af fokusmål. Vi vælger den variant af fokusordsmål, der korrelerer numerisk stærkest med læse- og stavemålene i korrelationsanalyserne. Da vores deltagergruppe er begrænset, kan vi ikke have alle de klassiske eye-tracking-mål med i regressionsanalyserne uden at det går ud over analysens mulighed for at vise signifikant bidrag fra de enkelte mål, så vi udvælger også her det mål, der korrelerer stærkest med læse- og stavemålene.

3 I de hierarkiske regressionsanalyser anvender vi læseforståelseseffektiviteten, altså antal korrekt besvarede spørgsmål per minut. Det skyldes, at vi ønsker at forklare den samlede læsefærdighed ud fra eye-tracking-målene.

Resultater

I dette afsnit belyses det, hvordan forbindelsen mellem læse- og stave-mål henholdsvis eye-tracking-mål kommer til udtryk, når eleverne i dette studie læser sammenhængende tekster med eye-tracker.

Som baggrund for korrelations- og hierarkiske analyser viser Tabel 2 deskriptiv statistik for de inkluderede læse-, stave- og eye-tracking-mål.

Tabel 2.

Deskriptiv statistik for alle mål.

	M	median	SD	Min	Max
Læse- og skrivemål					
Rigtige per minut ^a	3,04	2,90	1,21	0,80	6,50
Samlet besvarelsestid ^a	14,73	13,70	5,93	6,80	43,60
Procent i korrekt besvaret ^a	88,62	91,00	9,92	57,0	100,0
Andel korrekt stavede ord ^b	57,57	61,00	23,03	3,00	100,00
Eye-tracking-mål					
Læsehastighed (ord per min.)	112,32	117,82	30,68	39,67	178,53
Fokusord (procent af læste ord)	5,37	2,98	5,95	0,45	26,23

Mediale fokusord (procent af læste ord)	3,14	1,51	3,95	0,18	17,68
Wrap-up fokusord (procent af læste ord)	2,32	1,50	2,13	0,23	9,29
Fokusord (procent af læsetid)	14,62	10,49	11,42	2,49	51,62
Mediale fokusord (procent af læste ord)	8,04	5,27	7,31	0,78	32,56
Wrap-up fokusord (procent af læsetid)	6,87	5,65	4,64	0,85	21,88
Gns. varighed af ord-fikseringer (sekunder)	0,27	0,27	0,07	0,11	0,53
Gns. varighed af første fiksering pr. ord (sekunder)	0,28	0,28	0,08	0,11	0,59
"First pass dwell time" (sekunder)	67,14	67,98	21,31	14,59	111,00
Overskimmede ord i "first pass" (procent af læste ord)	0,46	0,48	0,11	0,17	0,69

^a Læseforståelse, Hogrefe,

^b Stavning, Hogrefe

Vi rapporterer gennemsnittet, medianen, standardafvigelsen og spredningen for hvert mål.

Vi vurderer styrken af sammenhængen mellem læse- og stavemål og de klassiske eye-tracking-mål med Spearmans rangkorrelationer (se Tabel 3). Vi leder efter sammenhænge mellem de to typer af mål, der minder om, dem vi kender fra eksperimentelle setup.

Korrelationerne i Tabel 3 viser, at læse- og stavemålene er moderat til stærkt korreleret med læsehastighed, med signifikante korrelationskoefficienter på mellem 0,420 og 0,655 og konfidensintervaller mellem 0,051 og 0,886. Disse sammenhænge bekræfter, at eleverne under optagelsen af øjenbevægelser, som forventet, har brugt tiden på almindelig oplæsning. Der er dermed ikke grund til at tro, at eleverne oplevede situationen som noget andet end normal læsning, og vi kan tolke resultaterne i dette lys. Det er værd at bemærke at styrken af korrelationen mellem læsehastighed og læse- og stavemålene minder om den, vi ser mellem de forskellige varianter af eye-tracking-mål og læse- og stavemålene. Der er således ingen eye-tracking-mål, der er stærkere forbundet til læse- og stavemålene end det simple mål af læsehastighed.

Af Tabel 3 fremgår det, at sammenhængene mellem læseforståelses- og stavemålene og den gennemsnitlige varighed af ord-fikseringer, samt den gennemsnitlige varighed af første fiksering per ord, er moderate, med signifikante negative korrelationskoefficienter mellem -0,469 og -0,565. Sammenhængen mellem læse- og stavemålene og de to tekstbaserede mål, first pass dwell time og overskimmede ord i first pass, er svage, og alle korrelationer er for disse to eye-tracking-mål ikke-signifikante. I eksperimentelle studier har alle fire klassiske mål vist sig at have en sammenhæng med læsefærdigheder. Det er dog ikke det mønster, som vi finder, når vi måler øjenbevægelser i en skolehverdag. Her er det kun to af fire eye-tracking-mål (gennemsnitlig varighed af ord-fikseringer og gennemsnitlig varighed af første fiksering per ord), der er robuste for den ekstra målestøj, der er forbundet med indsamling af øjenbevægelsesdata i et skolevenligt-setup.

Tabel 3.

Spearmans rho (r_s) korrelationskoefficienter for læse-, stave- og eye-tracking-mål.

	Læsning		Stavning	
	Rigtige per minut	Samlet besvarelsestid	Procent korrekt besvaret	Andel korrekt stavede ord
Eye-tracking-mål				
Læsehastighed (ord per min.)	0,615** (0,325; 0,807)	-0,592** (-0,781; -0,298)	0,420** (0,051; 0,737)	0,655** (0,320; 0,886)
Fokusord (procent af læste ord) A	-0,565** (-0,789; -0,261)	0,547** (0,217; 0,759)	-0,385** (-0,717; -0,031)	-0,607** (-0,818; -0,271)
Mediale fokusord (procent af læste ord) B	-0,526** (-0,766; -0,210)	0,493** (0,161; 0,741)	-0,385** (-0,725; -0,017)	-0,631** (-0,864; -0,309)
Wrap-up fokusord (procent af læste ord) B	-0,491** (-0,763; -0,143)	0,498** (0,148; 0,752)	-0,273 (-0,630; 0,113)	-0,487** (-0,735; -0,166)
Fokusord (procent af læsetid) A	-0,632** (-0,812; -0,362)	0,608** (0,305; 0,801)	-0,422** (-0,714; -0,053)	-0,629** (-0,805; -0,281)

Mediale fokusord (procent af læsetid) B	-0,562** (-0,777; -0,255)	0,520** (0,200; 0,756)	-0,440** (-0,743; -0,084)	-0,666** (-0,853; -0,332)
Wrap-up fokusord (procent af læsetid) B	-0,553** (-0,755; -0,211)	0,565** (0,239; 0,758)	-0,271 (-0,594; 0,134)	-0,469** (-0,748; -0,085)
Gns. varighed af ord-fikseringer	-0,469** (-0,770; -0,066)	0,465** (0,078; 0,754)	-0,319 (-0,703; 0,025)	-0,562** (-0,826; -0,267)
Gns. varighed af første fiksering pr. ord	-0,480** (-0,807; -0,102)	0,481** (0,121; 0,793)	-0,318 (-0,722; 0,006)	-0,565** (-0,817; -0,281)
"First pass dwell time"	-0,051 (-0,435; 0,412)	0,019 (-0,457; 0,430)	-0,132 (-0,496; 0,255)	-0,092 (-0,412; 0,283)
Overskimmede ord i "first pass" (procent af læste ord)	-0,155 (-0,484; 0,256)	0,085 (-0,357; 0,436)	-0,266 (-0,641; 0,164)	-0,216 (-0,543; 0,198)

n=68; A: n=65, B: n=64, ** angiver at korrelationen er signifikant $p < 0,05$ (Bonferroni-justeret to-halet). Tal i parentes er grænserne for 95%-konfidensintervallet (bootstrap-samplet med replacement, n=1000).

For at validere, at det nye eye-tracking-mål fokusord både er forbundet til læse- og stavfærdigheder og klassiske eye-tracking-mål, bruger vi Spearmans rangkorrelationer (se Tabel 3 og 4) til at vurdere styrken af sammenhængen mellem:

1. Klassiske eye-tracking-mål og det nye eye-tracking-mål fokusord (Tabel 4)
2. Læse- og stavemålene og det nye eye-tracking-mål fokusord (Tabel 3).

Spearmans rho korrelationerne i Tabel 4 viser, at der er stærke sammenhænge mellem fokusord og både den gennemsnitlige varighed af ord-fikseringer og den gennemsnitlige varighed af første fiksering per ord (r_s mellem 0,688 og 0,816 og konfidensintervaller der ligger mellem 0,369 og 0,936, se Tabel 4). Dette tolkes som et tegn på samstemmende validitet mellem fokusord og de to fikseringsmål. Omvendt er sammenhængen mellem fokusord og målene "First pass dwell time" og overskimmede ord i "first pass" ikke stærke (r_s mellem 0,145 og 0,425 og konfidensintervaller mellem -0,349 og 0,650 hvoraf kun et enkelt netop ikke inkluderer 0, se Tabel 4).

Spearmans rho-korrelationerne i Tabel 3 viser, at styrken af sammenhængen mellem læse- og stavemål og fokusord varierer med hver type af mål.

Sammenhængen mellem andel korrekt stavede ord er generelt stærkt forbundet med det samlede fokusordsmål og med mediale fokusord (r_s mellem $-0,666$ og $-0,599$, og konfidensintervaller mellem $-0,864$ og $-0,271$, se Tabel 3), og moderat forbundet til wrap-up fokusord (r_s på hhv $-0,487$ og $-0,469$, og konfidensintervaller mellem $-0,748$ og $-0,085$, se Tabel 3). De stærke til moderate sammenhænge tolkes, som et tegn på at fokusord, ligesom de klassiske eye-tracking-mål, er forbundet med stavefærdighed.

Sammenhængen mellem fokusordsmålene og læsemålene er generelt moderate. Kun for læsemålet procent korrekt besvarede læseforståelsesspørgsmål finder vi svage og ikke signifikante sammenhænge, nemlig med wrap-up-varianten af fokusordsmålet. Det tolker vi som et udtryk for, at fokusord særligt er forbundet med samlet læsefærdighed, og i mindre grad med læseforståelsespræcision. Derfor anvender vi det bredere læseforståelseseffektivitetsmål i de hierarkiske regressionsanalyser (se Tabel 6).

I de hierarkiske regressionsanalyser (se Tabel 5 og 6) vil vi gerne anvende det samme fokusordsmål, uanset om det er stavning eller læseforståelseeffektivitet, vi ønsker at forklare variation i. Fokusord opgjort som procent af læsetid opnår med en enkelt undtagelse konsistent numerisk højere signifikante Spearmans rho-korrelationer (se Tabel 3) med læse-/stavemålene end fokusord opgjort som procent af læste ord. Vi vælger mediale fokusord, da det mål har den højeste observerede korrelation med stave- og læsemålene ($r_s = -0,666$ og konfidensintervallet $-0,853$ til $-0,332$) og samtidig udviser stærk sammenhæng på tværs af stave- og læsemål.

Tabel 4.

Spearmans rho korrelationskoefficienter for klassiske eye-tracking-mål og fokusord.

	Klassiske eye-tracking-mål			
	Gns. varighed af ord-fikseringer	Gns. varighed af første fiksering pr. ord	"First pass dwell time"	Overskimmede ord i "first pass" (procent af læste ord)
Fokusord				
Fokusord (procent af læste ord) A	0,811** (0,577; 0,935)	0,816** (0,587; 0,936)	0,264 (-0,206; 0,633)	0,380 (-0,050; 0,706)
Mediale fokusord (procent af læste ord) B	0,770** (0,502; 0,907)	0,784** (0,523; 0,922)	0,220 (-0,261; 0,560)	0,344 (-0,073; 0,628)
Wrap-up fokusord (procent af læste ord) B	0,816** (0,543; 0,931)	0,814** (0,558; 0,930)	0,331 (-0,105; 0,650)	0,425** (0,008; 0,752)
Fokusord (procent af læsetid) A	0,753** (0,474; 0,908)	0,752** (0,480; 0,915)	0,190 (-0,318; 0,594)	0,312 (-0,155; 0,630)
Mediale fokusord (procent af læsetid) B	0,707** (0,413; 0,896)	0,715** (0,402; 0,902)	0,145 (-0,349; 0,541)	0,278 (-0,147; 0,601)
Wrap-up fokusord (procent af læsetid) B	0,701** (0,392; 0,885)	0,688** (0,369; 0,883)	0,194 (-0,259; 0,573)	0,299 (-0,184; 0,631)

n=68, A. n=65, B. n=64, ** angiver at korrelationen er signifikant $p < ,05$ (Bonferoni-justeret to-halet). Tal i parentes er grænserne for 95%-konfidensintervallet (bootstrap-samplet med replacement, n=1000).

Vi laver to multiple hierarkiske regressionsanalyser i to trin for at forklare variation i henholdsvis læseforståelseeffektivitet og andel korrekt stavede ord. I første trin inddrager vi det klassiske eye-tracking-mål gennemsnitlig varighed af første fiksering pr. ord., da dette mål er det af de fire klassiske eye-tracking-mål, der har den numerisk stærkeste sammenhæng med de to læse- og stavemål (Tabel 3). Det klassiske mål inddrages i trin 1. I trin 2 inddrages målet mediale fokusord

(procent af læsetid), som er det af de seks fokusordsmål vi vurderer, er bedst til at forklare variation i stavning og læsning på tværs af målene andel korrekt stavede ord og læseforståelseseffektivitet (Tabel 3).

Tabel 5 viser modellerne med detaljer for andel korrekt stavede ord og Tabel 6 for læseforståelseseffektivitet. Antagelserne for den multiple hierarkiske regressionsanalyse er i vidt omfang overholdt (Se Bilag 2).

Tabel 5.

Multipl hierarkisk regression.

Variabel	Andel korrekt stavede ord			
	Model 1		Model 2	
	B	β	B	β
Konstant	100,553**		70,475**	
Gns. varighed af første fiksering pr. ord	-152,703**	-,491	18,328	,059
Mediale fokusord (procent af læsetid)	-	-	-2,431**	-0,745
R ²	,241		,478	
F	19,335**		29,347**	
ΔR^2	,241**		,254**	
ΔF	19,335**		30,127**	

n=63, **p<,001. Note. Niveauet for signifikans er bonferroni justeret til p=,025 for to outcomevariable i samme datasæt. Forklaret variation i andel korrekt stavede ord fra gns. varighed af første fiksering pr. ord og mediale fokusord.

I modellen med antal af korrekt stavede ord som outcome-mål gjaldt det, at den fulde model, som bruger gennemsnitlig varighed af første fiksering pr. ord og mediale fokusord til at forudsige andel korrekt stavede ord, var signifikant (tilpasset R² =,478, F(2, 60) =29,347, p <,001, se Tabel 5).

Tilføjjelsen af mediale fokusord som procent af læsetid til forklaringen af andel korrekt stavede ord (Model 2 i Tabel 5) ledte til en signifikant forøggelse af R^2 på $\Delta R^2 = ,254$, $F(1, 60) = 30,127$, $p < ,001$. For deltagerne i dette studie forklarer det nye fokusordsmål altså en unik andel af

Tabel 6.

Multipel hierarkisk regression .

Variabel	Læseforståelseseffektivitet			
	Model 1		Model 2	
	B	β	B	β
Konstant	4,816**		3,524**	
Gns. varighed af første fiksering pr. ord	-6,301*	-,392	,987	,061
Mediale fokusord (procent af læsetid)	-	-	-,101**	-,609
			,320	
R^2	,154			
F	11,269*		14,855**	
ΔR^2	,154*		,166**	
ΔF	11,269*		14,855**	

variationen i andel korrekt stavede ord.

$n=64$, * $p<,001$, ** $p<,001$. Note. Niveauet for signifikans er bonferroni justeret til $p=,025$ for to outcomevariable i samme dataseæt.

Forklaret variation i læseforståelseseffektivitet fra gns. varighed af første fiksering pr. ord og mediale fokusord.

I modellen med læseforståelseseffektivitet som outcome-mål gjaldt det, at den fulde model, som bruger gennemsnitlig varighed af første

fiksering pr. ord og mediale fokusord til at forudsige læseforståelses-effektivitet, var signifikant (tilpasset $R^2 = ,320$, $F(2, 61) = 14,855$, $p < ,001$ se Tabel 6). Tilføjelsen af mediale fokusord som procent af læsetid til at forklare læseforståelseeffektivitet (Model 2 i Tabel 6) ledte til en signifikant forøgelse af R^2 på $\Delta R^2 = ,166$, $F(2, 62) = 14,855$, $p < ,001$.

I de fulde modeller for læseeffektivitet (Tabel 6) og andel korrekt stavede ord (Tabel 5) er det kun den standardiserede beta-koefficient for mediale fokusord, der er signifikant forskellige fra nul, hvilket indikerer, at denne variabel er lineært forbundet til læseforståelses-effektivitet og stavning. Desto mere tid eleverne bruger på fokusord i medial position i sætningen, desto dårligere læseforståelse har de, og desto dårligere staver de. De standardiserede beta-koefficienter viser, at én standardafvigelses længere tid brugt på fokusord i medial position, hænger sammen med en lavere score i andel korrekt stavede ord svarende til $-0,745$ standardafvigelser og en lavere læseforståelse-effektivitet svarende til $-0,609$ standardafvigelser. I afrundede tal betyder det at hver yderligere 4-5 sekunder ud af et minuts læsning brugt på mediale fokusord svarer til et forventet fald på næsten to færre korrekt stavede ord og trekvart færre korrekt besvarede læseforståelsesspørgsmål per minut i de standardiserede test.

Vi har eksplorativt lavet endnu et sæt hierarkiske regressionsanalyser, da vi gerne vil belyse om mediale fokusord forklarer variation i henholdsvis læseforståelseeffektivitet og stavepræcision ud over simpel læsehastighed (se bilag 3). Vi kan se af de analyser, at mediale fokusord ikke forklarer unik variation i læseforståelseeffektivitet (tabel 8, bilag 3) ud over et simpel mål af læsehastighed, men at tilføjelsen af mediale fokusord til læsehastighed til at forklare præcision i stavning (Model 2 i Tabel 7) ledte til en signifikant forøgelse af R^2 på $\Delta R^2 = ,093$, $F(2, 65) = 9,801$, $p < ,01$.

Diskussion og konklusion

Vi har undersøgt, hvorvidt eye-tracking-mål i et skolevenligt-setup afspejler elevers læse- og stavefærdigheder, ved at optage 68 elevers øjenbevægelser under læsning og sammenholde dem statistisk med elevernes resultater på standardiserede læse- og stavetest. Som svar på det første forskningsspørgsmål, så understøtter undersøgelsens resultater, at begge de klassiske eye-tracking-mål, der er

defineret af fikseringstider, er forbundet med læse- og stavemålene, når elever læser sammenhængende tekster med eye-tracker i et skolevenligt setup. Det betyder, at disse klassiske eye-tracking-mål er robuste nok til at blive indsamlet i et skolevenligt setup og stadig afspejle elevernes læse- og stavefærdigheder. Det har begrænset direkte praktisk relevans da disse mål er svære at tolke, men det tillader os at anvende dem som grundlag for valideringen af det nye fokusordsmål i dette studie.

Fra eksperimentelle studier ved vi, at den samlede læseproces bidrager til fikseringstiden, som derfor vil være påvirket af først afkodnings- og siden læseforståelsesprocesser. I det skolevenlige-setup finder vi også, at målet har sammenhæng med både stave- og læseforståelsesmålene.

Varigheden af første fiksering pr. ord sammenkædes i litteraturens mere eksperimentelle forsøgsdesign tættest med afkodningshastighed end forståelsesprocesser. Afkodning og stavning er, som tidligere beskrevet, tæt forbundne færdigheder. Og selvom stavning stiller endnu højere krav til præcisionen i de lagrede ortografiske repræsentationer, end afkodning gør, er stavemålet undersøgelsens bedste korrelat til stærkeste mål for elevernes afkodningsfærdighed. Vi finder også, som forventet, at første fiksering pr. ord er stærkest korreleret med stavepræcisionsmålet, men både distributionen (Tabel 2) og korrelationerne (Tabel 3) for første fikserings varighed er stort set identiske med målet for den gennemsnitlige fikseringstid. I gennemsnit blev hvert ord fikseret 1,75 gange, altså under to, hvilket betyder, at de fleste ord kun fikseres én gang, og indikerer at de første fikseringer er dominerende i datasættet. Dette afspejler også, at teksten har relativt mange korte, almindelige ord, og at den er let at læse for denne elevgruppe.

Vi ser altså, at elever der staver sikkert, er tilbøjelige til at have kortere varighed af første fiksering pr. ord og kortere gennemsnitlig fikseringstid, mens elever, som staver mindre præcist, er tilbøjelige til at have længere varighed af første fiksering pr. ord og længere gennemsnitlig fikseringstid. Vi tolker dette som et udtryk for, at øjenbevægelserne påvirkes systematisk af den underliggende variation i deltagerens afkodnings- og stavefærdighed.

De klassiske eye-tracking-mål udviser en robust sammenhæng med læse- og stavemål i undersøgelsens skolevenlige-setup. Den mest oplagte forklaring er, at vi i denne undersøgelse, ligesom for standardiserede læse- og stavemål, sammenligner alle elevs læsning af samme tekst. Det er sandsynligt at de klassiske eye-tracking-mål ville have en svagere sammenhæng med læse- og stavemål, hvis eleverne havde læst forskellige tekster. I et design med forskellige tekster er det vores hypotese, at variationen i øjenbevægelser, der skyldes stave- og læse-

færdigheder, og som dermed driver sammenhængen med de standardiserede læse- og stavemål, kan blive overskygget af den variation i øjenbevægelser, som skyldes teksternes forskellighed, fx. i ord- og sætningslængde. Denne undersøgelses design med én fælles tekst kan derfor let overvurdere de klassiske eye-tracking-måls sammenhæng med læse- og stavemål i et skolevenligt-setup.

De tekstorienterede eye-tracking-mål, first pass dwell time og andel overskimmede ord var mindst følsomme for forskelle i læse- og stavemålene i vores studie. Dette er et overraskende resultat, da vi forventede, at bedre læsere ville score lavere på det første mål, og højere på det sidste mål, fordi læsere med sikker afkodning og god forståelse bedre kan udnytte skimmestrategier og visuel perifer information. Målenes ringe evne til at afspejle elevernes læse- og stavefærdigheder kan skyldes, at måden teksten blev præsenteret på, kan have forurennet begge mål. Det kan være sket, fordi læserens blik ofte slutter med at fokusere på bladreknapen nederst til højre og derefter søger diagonalt op over den nye side for at læse videre. Da begge mål forudsætter, at den første fiksering på et ord på siden er et forsøg på at læse ordet i kontekst, regnes alle foranstående ord på samme side, der endnu ikke har været fikseret, som oversprungne. Derfor kan korte orienterende blik, uden reel betydning for læsningen, have undergravet disse to måls validitet i et skolevenligt setup. Den forklaring kan testes ved at annotere øjenbevægelsesdata manuelt, så eventuelle orienteringsblik skæres fra på hver læst side, men vi har valgt ikke at gå videre med at undersøge muligheden, da den type manuel annotering ikke er tilgængelig i skolen i praksis. Ekstra blanke sider med fikseringskryds mellem hver bladring kunne eliminere den risiko, men de påvirker læserens arbejdshukommelse og flow uhensigtsmæssigt, og gør samtidig læsning med eye-tracker påfaldende anderledes og mindre attraktivt didaktisk.

Vi finder, som svar på det andet forskningsspørgsmål, at det nye fokusordsmål korrelerer med læse- og stavemål mindst i samme grad som de klassiske eye-tracking-mål, det teoretisk er sammenligneligt med, og som vi derfor har udvalgt i denne undersøgelse. Vi har undersøgt fokusord både målt som andelen af læste ord og andelen af læsetid brugt på fokusordene. Det generelle mønster i korrelationsanalysen er, at andelen af tid brugt på fokusord er stærkest korreleret med undersøgelsens læse- og stavemål. Det peger på, at længere afbrydelser af den flydende læsning kan være et lidt tydeligere tegn på læse- eller stavevanskeligheder end flere afbrydelser.

Vi har isoleret fokusord der står ved en sætningsgrænse, og fokusord der ikke gør, fordi opbremsninger ved sætningsgrænser er almindelige, også for stærke læsere, og ofte lyder upåfaldende under oplæsning. Korrelationsanalysen afspejler dette, og vi finder, at de mediale

fokusord er konsistent stærkere associeret med stavemålet og det læseforståelsesmål som vægter præcision frem for hastighed, sammenlignet med wrapup-varianten. Det indikerer, at opbremsninger midt i sætninger er et tydeligere tegn på vanskeligheder med præcision i læsning og stavning end opbremsninger ved sætningsgrænser.

I korrelationsanalyserne i denne undersøgelse er fokusordsmålet stærkest forbundet til læse- og stavfærdighed, når det defineres som andelen af tid brugt på mediale fokusord, men spredningen og overlappene i konfidensintervallerne betyder at en gentagelse af studiet kunne falde ud til fordel for en anden variant af fokusordsmålet.

Som svar på tredje forskningsspørgsmål, viser de hierarkiske regressionsanalyser, at fokusord forklarer unik variation både i elevernes læse- og stavfærdighed ud over den variation, der forklares af det traditionelle eye-tracking-mål, der er stærkest korreleret med læse- og stavemålene. Resultaterne underbygger, at målet fokusord, der er udformet med praksis for øje, på samme måde som klassiske eye-tracking-mål, kan afspejle elevernes læse- og stavfærdigheder, og endda forklarer variation i disse ud over den, som de klassiske mål kan indfange. Det kan skyldes, at fokusord er designet til at detektere fluency-nedbrud, hvilket betyder, at det er følsomt for nedbrud i både afkodning og forståelse. Det er netop en væsentlig del af, hvad standardiserede stave-, afkodnings- og læseforståelsestest måler, hvorimod klassiske eye-tracking-mål især er brugt i læseforskning til at skelne variation i trænede læsers flydende læsning. Vi kan dog ikke ud fra denne undersøgelse forklare hvad der præcis gør, at fokusord bidrager til at forklare unik variation i læse- og stavfærdighed ud over den et klassisk eye-tracking-mål forklarer.

Vi finder samtidig at simpel læsehastighed er bedre til at forklare variation i læseforståelseseffektivitet end fokusord. Det resultat underbygger, at fokusords bidrag til at forklare forskelle i læseforståelseseffektivitet i høj grad knytter sig til læsehastigheden. Den variation fokusord forklarer er indeholdt i den variation læsehastighed forklarer. Vi bliver på den måde ikke bedre til at forklare forskellene i elevernes læseforståelseseffektivitet gennem måling af fokusord end gennem måling af simple læsehastighed. Dette er dog ikke tilfældet for stavning, hvor fokusord forklarer unik variation ud over den som forklares af læsehastigheden. Fokusord bidrager ud over læsehastighed, når vi skal forklare stavning. Det resultat kan tale for, at det særligt er følsomheden for fluency-nedbrud, der knytter sig til afkodning, der vægter i denne undersøgelse af fokusord. I det lys kan det særligt væsentlige bidrag fra fokusord være at udpege ord, som har givet anledning til fluency-nedbrud i afkodningen for den enkelte læser. Hvis senere undersøgelser kan skabe klarhed over i hvor høj grad fokusord for den enkelte elev modsvarer ord eleven, havde svært ved at læse,

så følger der muligheder for at anvende fokusord som baggrund for undervisningens tilrettelæggelse.

I denne undersøgelse viser vi, at en enkelt kort læsning i en letlæst bog kan afspejle relevante forskelle i elevernes stave- og læsefærdigheder på fjerde klassetrin på to skoler i to forskellige kommuner. Resultaterne kan dog ikke frit generaliseres til andre elevgrupper. Fremtidige studier kan med en stikprøve, der er repræsentativ for danske skoleelever på flere forskellige alderstrin og læse- og staveniveauer, svare på, om resultaterne kan generaliseres til flere elevgrupper

Perspektiver for praksis

Denne undersøgelse adresserer to nødvendige skridt på vejen mod praktisk og metodisk modenhed for eye-tracking-teknologien som et redskab, der gennem optagelse af elevernes øjenbevægelser ved læsning af selvvalgte tekster, kan generere viden, som lærere og læsevejledere kan anvende i arbejdet med at tilpasse læseundervisningen til den enkelte elev.

Det første skridt er at afdække, om forskelle mellem eleverne i øjenbevægelser optaget under selvstændig hverdagslæsning faktisk afspejler forskelle i elevernes læseforståelses- og stavefærdighed. De moderate til stærke sammenhænge mellem klassiske eye-tracking-mål, fokusordsmålene og målene af henholdsvis læseforståelse og stavning underbygger dette. Resultatet bidrager dermed til dette første skridt på vejen mod teknologiens praktiske og metodiske modenhed. Samtidig er det også væsentligt at fremhæve at eleverne i denne undersøgelse alle læste samme tekster. Vi ved derfor endnu ikke om sammenhængen mellem læseforståelse, stavning og fokusord ville være mindre i andre situationer. Denne undersøgelse er endvidere en her-og-nu undersøgelse. Vi kan derfor ikke belyse, om optagelser af øjenbevægelser reflekterer elevernes læse- og staveudvikling over tid.

Et spørgsmål i relation til dette første skridt kan fx være, om elevernes læse- og staveudvikling kan følges, i en ramme der svarer til almindelig frilæsning, mellem tidspunkterne for de standardiserede test, og dermed hjælpe skolen med løbende at opdage og hjælpe elever, der har brug for særlig støtte. Resultaterne, præsenteret her, er baseret på, at eleverne læser samme tekst i et begrænset tidsrum. Med designet lægger undersøgelsen sig tæt op af den ramme som forudsættes under standardiserede tests, og giver dermed så præcist et sammenligningsgrundlag som muligt mellem øjenbevægelsesmåling

og standardiserede mål for læsning og stavning, der er almindeligt tilgængelige i en skolekontekst. En oplagt fortsættelse er at undersøge, hvordan øjenbevægelsesmål kan bruges med elever, der læser forskellige tekster.

Det næste skridt mod teknologiens praktiske modenhed i en skolekontekst, er at udvikle mål, der kan tolkes og anvendes med den eksisterende specialviden, som læsevejlederen allerede besidder. Fokusordsmålet er et bud på sådan et mål. I denne undersøgelse finder vi, at fokusord er et robust mål. Det underbygges dels af de moderate til stærke korrelationer med læseforståelse og stavning, og dels af at fokusord forklarer variation i læse- og stavemål ud over den, som de klassiske mål kan forklare. Undersøgelsen belyser fokusords sammenhæng med læse- og stavemål på gruppeniveau. Videre undersøgelser, der sammenligner analyser af elevers højt læsning og de udpegede fokusord på individniveau, kan besvare i hvor høj grad fokusord udpeger ord, som den enkelte elev har svært ved at læse. Denne videre analyse er nødvendig for at belyse, hvordan udpegningen af fokusord kan anvendes i lærerens og læsevejlederens arbejde med at tilpasse læseundervisningen til den enkelte elev.

På den baggrund tager denne undersøgelse et skridt på vejen mod et længere proces omkring at validere brugen af eye-tracking, som en måde at få relevant viden om elevernes skriftsproglige udvikling i et naturligt skole-setup, og som en mulighed for at få viden om den enkelte elevs læsninger, som baggrund for at tilrettelægge undervisningen.

Referencer

- Alexander-Passe, N.** (2008). The sources and manifestations of stress amongst school-aged dyslexics, compared with sibling controls. *Dyslexia: An International Journal of Research and Practice*, 14(4), 291-313. <https://doi-org.ez-ucs.statsbiblioteket.dk/10.1002/dys.351>
- Andersen, D.** (2005). 4 år efter grundskolen – 19-årige om valg og veje i ungdomsuddannelserne. København: AKF-forlaget
- Benfatto, M. N., Seimyr, G. Ö., Ygge, J., Pansell, T., Rydberg, A. & Jacobson, C.** (2016). Screening for dyslexia using eye tracking during reading. *PLoS ONE*, 11(12), 1-16. <https://doi.org/10.1371/journal.pone.0165508>
- Bundsgaard, J. & Puck, M. R.** (2016). Nationale Test – Danske Lærere Og Skolelederes Brug, Holdninger Og Viden. DPU, Aarhus Universitet.

- Børne- og undervisningsministeriet** (2019). *Dansk Fælles Mål*. Lokaliseret d. 12. januar 2023 på https://emu.dk/sites/default/files/2020-09/GSK_F%C3%A6llesM%C3%A5l_Dansk.pdf
- Børne- og undervisningsministeriet** (2022). *Folkeskolens Nationale Færdighedstest*. Lokaliseret d. 12. januar 2023 på <https://www.uvm.dk/folkeskolen/test-evaluering-og-skoleudvikling/test--og-evalueringsredskaber/folkeskolens-nationale-faerdighedstest>
- Caravolas, M., Hulme, C. & Snowling, M.** (2001). The Foundations of Spelling Ability: Evidence from a 3 Year Longitudinal Study. *Journal of Memory and Language - J MEM LANG*, 45, 751-774.
- Carroll J. M. & Iles J. E.** (2006). *An assessment of anxiety levels in dyslexic students in higher education*. *British Journal of Educational Psychology*, 76, 651-662.
- Christensen, V. T.** (2019). *PISA 2018 – Danske unge i en international sammenligning*. København: VIVE. Det Nationale Forsknings-og Analysecenter for Velfærd.
- Elbro, C.** (2021). *Læsevanskeligheder*. Hans Reitzels Forlag. København. ISBN: 9788741263359
- Elbro, C.** (2014). *Læsning og læseundervisning*. Hans Reitzels Forlag. København. ISBN: 9788762501799
- Franzen, L., Stark, Z. & Johnson, A. P.** (2021). Individuals with dyslexia use a different visual sampling strategy to read text. *Scientific reports*, 11(1), 6449.
- Gran Ekstrand, A. C., Nilsson Benfatto, M. & Öqvist Seimyr, G.** (2021, March). Screening for Reading Difficulties: Comparing Eye Tracking Outcomes to Neuropsychological Assessments. In *Frontiers in Education* (Vol. 6, p. 643232). Frontiers Media SA.
- Hinkle, D. E., Wiersma, W. & Jurs, S. G.** (2003). *Applied Statistics for the Behavioral Sciences*. Boston, MA: Houghton Mifflin Company.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & Van de Weijer, J.** (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Juul, H.** (2019). *Vejledning til staveprøve 1-3*. Høgrete Psykologisk Forlag
- Kousholt, K.** (2015). "Børns Gættier Ved Nationale Test". *Cebra-sriben, Nationale Tests*, nr. 18: 46-57.
- Laerd Statistics** (2018). Spearman's correlation using SPSS Statistics. *Statistical tutorials and software guides*. <https://statistics.laerd.com/>
- Larsen, M., Jakobsen, V. & Rosdahl, A.** (2022). *Voksnes basale færdigheder. Udvikling i læse- og regnefærdigheder de seneste 10 år*. VIVE. e-ISBN: 978-87-7582-071-9
- Meister, C., Pimentel, T., Clark, T., Cotterell, R. & Levy, R.** (2022). Analyzing Wrap-Up Effects through an Information-Theoretic Lens. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 20-28).
- Møller, L.** (2013). *Vejledning til tekstlæseprøve 1-8*. Høgrete Psykologisk Forlag
- Nelson, J. M., Lindstrom, W. & Foels, P. A.** (2015). Test Anxiety Among College Students With Specific Reading Disability (Dyslexia): Nonverbal Ability and Working Memory as Predictors. *Journal of Learning Disabilities*, 48(4), 422-432. <https://doi.org/10.1177/0022219413507604>

- Ravn, K.** (2015). "Duer ikke: En femtedel af opgaverne i de nationale test kasseret – Folkeskolen.dk". folkeskolen.dk. <https://www.folkeskolen.dk/572751/duer-ikke-en-femtedel-af-opgaverne-i-de-nationale-testkasseret>
- Rayner, K.** (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.
- Rosdahl, A., Fridberg, T., Jakobsen, V. & Jørgensen, M.** (2013): "Færdigheder i læsning, regning og problemløsning med IT". København: SFI – Det Nationale Forskningscenter for Velfærd
- Rosdahl, A., Fridberg, T., Jakobsen, V. & Jørgensen, M.** (2013): "Færdigheder i læsning, regning og problemløsning med IT. Sammenfatning af resultater fra PIAAC". København: SFI – Det Nationale Forskningscenter for Velfærd.
- Sperlich, A., Schad, D. J. & Laubrock, J.** (2015). When preview information starts to matter: Development of the perceptual span in German beginning readers. *Journal of Cognitive Psychology*, 27(5), 511-530.
- Swalander, L.** (2012). Selvbillede, motivation og dysleksi. In *Samuelsson (Ed.) Dysleksi og andre vanskeligheder med skriftsproget* (pp. 178-191). Dansk Psykologisk Forlag.
- Von der Malsburg, T. & Vasishth, S.** (2011). What is the scanpath signature of syntactic reanalysis?. *Journal of Memory and Language*, 65(2), 109-127.
- Wangebo, K.** (2011), Running record en systematisk og detaljeret iagttagelse af læseadfærd. *Viden om Læsning*, 9, 23-30. Nationalt Videncenter for Læsning.

Bilag 1

Kriterier for øjendatakvalitet

I vores forsøg har vi tilstræbt at eleverne kunne gennemføre læsningerne med eye-tracking på egen hånd, og der kan derfor være opstået afvigelser under optagelsen som gør de registrerede øjendata uegnede til brug i denne undersøgelse. Det kan fx. ske hvis eleven er blevet forstyrret, har misforstået opgaven, hvis lyset har ændret sig markant eller trackeren er blevet flyttet undervejs. Vi har valgt kriterier for datakvalitet, der tager højde for en skolehverdag hvor en gennemlytning af hver optagelse ikke er realistisk. De følgende fire kriterier har afgjort hvilke optagelser der skulle inspiceres manuelt og evt. ekskluderes fra analysen:

— *Andel af fikserede ord er under 85%*

Dette forekommer fx. når læseren bladrer videre uden at læse i bogen og hvor eye-trackingen er blevet upålidelig.

— *Læsehastighed er målt til at ligge uden for intervallet 20-180 ord/minut*

Høj læsehastighed forekommer under almindelig stillelæsning mens lav hastighed måles hvis teksten er alt for svær, hvis der er foregået noget andet end læsning, eller hvor trackingen er blevet upålidelig.

- *Kalibreringspræcision er målt til lavere end 4 ud af 5 mulige*
Dette forekommer når kalibreringen af eye-trackeren fejler. Det kan skyldes at instruktionen ikke er fulgt eller at det billede som eye-trackeren optager af læserens øjne, ikke viser pupillen tydeligt nok.

Kriterierne førte til at 9 (11%) af optagelserne blev inspiceret manuelt. Alle 9 blev udelukket fra analysen.

Optagelser med upålidelige øjendata kan stadig afspilles med lyd og manuelt evalueres af læreren, men det er ikke validt at foretage beregninger baseret på upålidelige optagelser af øjenbevægelser.

Detaljeret beskrivelse af læse- og stavetest i undersøgelsen.

Tekstlæseprøve 5 (Møller, 2013): Tekstens karakteristika er beskrevet med antal ord i teksten (946), ord pr side (59), sider (19), ord med mere end seks bogstaver (97), andelen af lange ord (10,2), sætninger (113), gennemsnitligt antal ord pr sætning (8,3) og lix (19). Der er ikke en tidsgrænse på testen, men eleverne opmuntres til, at løse så mange opgaver de kan på 20 minutter. Elever, der bruger mere tid, får lov at fortsætte. Den samlede besvarelestid måles. Testen beregner 1) antal korrekt besvarede per minut, 2) hastighed = antal forkert + korrekt besvarede per minut, 3) præcision = korrekt besvarede / (korrekt + forkert besvaret), 4) passerede (korrekt + forkert + oversprungne) og 5) antal oversprungne.

Staveprøve 3 (Juul, 2019): Items belyser om staveren staver lydligt acceptabelt, kan bruge hyppige betingede stavemåder, grammatiske suffikser og ordspecifikke stavemåder, samt mestrer stavning af sammensatte ord. Testen er ikke tidsbestemt. Testen beregner både andel korrekt stavede og lydligt acceptabelt stavede ord. Den beregner også andel korrekt stavede for betingede stavemønstre (herunder sammensatte ord), grammatiske suffikser og ordspecifikke stavemåder og stavning af sammensatte ord.

Statistisk kodning af nul fokusord

I dette valideringsstudie har vi valgt at udelukke 8 læsninger med nul fokusord for at opnå et datasæt med mindst mulig støj til validering. Nul Fokusord er den mest støjfyldte score fordi den ville tælle som den bedste score, og samtidig have stor risiko for at fange eye-tracking med mange udfald, uopmærksom hurtiglæsning uden forståelse og stillelæsning. Uden manuel inspektion, må scoren derfor antages at være tvetydig.

Scoren 0 er også problematisk i rangkorrelationsanalysen da den vil skabe et loft i den ene ende af skalaen, med atypiske ties for de tidsbaserede varianter af fokusord. Ved at udelade den potentielt tvetydige 0-score og dermed regne læsninger med 1 fokusord, besøgt i 1 sekund som optimalt, kan vi pålideligt vurdere forskellen på at beregne fokusordsprocent af tiden og af antallet af ord, samt forskellen på fokusordenes placering i sætningen.

I praksis vil en lærer ved manuel gennemlytning og inspektion af optagelsens kvalitet kunne vurdere om resultatet nul fokusord er pålideligt.

Detaljer om optagelse med eye-tracker

Trackeren blev placeret med magnet under skærmen, og den blev kalibreret i ni punkter inden optagelserne.

Fikseringer er beregnet på gennemsnittet af højre og venstre øje med I-VT-algoritmen (Komogortsev et al., 2010), med en nedre grænse for fikseringstid på 50 ms og en saccade-hastighedsparameter på $0,5^\circ/s$.

Bilag 2

Detaljeret beskrivelse af antagelser bag de Hierarkiske regressionsanalyser

(Tabel 5 og 6)

Antagelserne for den multiple hierarkiske regressionsanalyse for forudsigelsen af stavning er i høj grad overholdt, i det der var en rimelig lineær sammenhæng mellem outcome-variablen og hver af de to prædiktorvariable for sig og sammen. Dette vurderede vi på baggrund af visuel inspektion af de delvise regressionsplot og studentized residualer-plottet mod de forudsagte værdier. Analysen havde en Durbin Watson-værdi på 1,85, hvilket ikke indikerede, at de enkelte observationer var forbundet til hinanden, hvorfor residualerne vurderes at være uafhængige. Vi vurderede på baggrund af visuel inspektion af studentized residualer-plottet mod de forudsagte værdier, at der var varianshomogenitet for stavning. Toleranceværdien på 0,46, indikere-

de ikke multikollinearitet for nogle af variablene, hvorfor den stærke sammenhæng mellem gennemsnitlig varighed af første fiksering pr. ord og mediale fokusord (procent af læsetid) vurderes til at kunne indgå i samme analyse uden at overtræde antagelsen om multikollinearitet. For stavning er der en enkelt ekstrem værdi med studentized slettet residual på -3,66. Analysen af særligt indflydelsesrige værdier viste, at ingen leverage-værdier var større end 0,27, hvilket er acceptabelt, men samtidig pegede på to indflydelsesrige datapunkter med værdier for Cooks afstand på 0,15 og 0,12. En gennemgang af data indikerer ikke at disse datapunkter er fejl og vi beholder dem i analysen. Sidst vurderede vi på baggrund af visuel inspektion af et PP-Plot, at antagelsen om normalfordelte residualer blev overholdt.

For forudsigelsen af læseforståelseseffektivitet er næsten alle antagelser overholdt. Det gælder dog ikke antagelsen om varianshomogenitet, og der er endvidere et indflydelsesrigt datapunkter med værdier for Cooks afstand på 0,36. Vi har fastholdt analysemetoden dette til trods, da vi vurderer det som væsentligt, at de to outcome-variable læsning og stavning belyses med samme analyse.

Bilag 3

Analyserne i tabel 7 og 8 er gennemført som en eksplorative analyser efter designet af undersøgelsen. Formålet med analyserne er at belyse bidraget fra eye-tracking-mål i sammenligning med simpel læsehastighed. På den måde vil vi imødegå, at vi med valideringen af det nye eye-tracking-mål op ad eksisterende eye-tracking-mål kommer til, at fremstille resultaterne uden at læseren har en fornemmelse af, om det nye eye-tracking-mål forklarer forskelle i læse- og stavfærdighed, som rækker ud over det velkendte bidrag fra simpel læsehastighed. Læsehastighed er i denne undersøgelse opgjort som antal læste ord per minut.

Tabel 7.

Multipel hierarkisk regression.

Variabel	Andel korrekt stavede ord			
	Model 1		Model 2	
	B	β	B	β
Konstant	4,443		38,977	
Læsehastighed (ord per minut)	,517**	,539	,275(*)	,286
Mediale fokusord (procent af læsetid)	-	-	-2,586*	-,396
R ²	,290		,383	
F	27,011**		20,208**	
ΔR^2	,290**		,093*	
ΔF	27,011**		9,801*	

n=67, (*)p=,027 *p<,01 **p<,001. Note. Niveaueet for signifikans er bonferroni justeret til p=,025 for to outcomevariable i samme datasæt.
 Forklaret variation i andel korrekt stavede ord fra læsehastighed og mediale fokusord.

Tabel 8.

Multipel hierarkisk regression.

Variabel	Læseforståelseseffektivitet			
	Model 1		Model 2	
	B	β	B	β
Konstant	-,187		,443	
Læsehastighed (ord per minut)	,032**	,684	,028**	,590
Mediale fokusord (procent af læsetid)	-	-	-,045	-
				,146
R ²	,468		,480	
F	58,831**		30,451**	
ΔR^2	,468**		,012	
ΔF	58,831**		1,570	

n=68, *p=,001, **p<,001. Note. Niveauet for signifikans er bonferroni justeret til p=,025 for to outcomevariable i samme datasæt. Forklaret variation i læseforståelseseffektivitet fra læsehastighed og mediale fokusord.