**Dokumentation af Sætningslæseprøver**

**Peter Allerup januar 2014**

**Repræsentativitet af data**

Data blev indsamlet til forårets runde som en simpel tilfældig stikprøve bestående af 31 skoler, udtaget fra en bruttoliste af ca. 80 skoler, der reagerede positivt på en generel invitation til at deltage i afprøvningen af de nye læseprøver fra Hogrefe. Fra disse 31 skoler blev der udtaget klassetrin, som er relevante for læseprøverne. Alle klasser på de udvalgte klassetrin blev inviteret til at deltage. Det valgte grundlag med de 80 skoler medfører, at skolerne *ikke* er et tilfældigt sample og den statistiske analyse herunder skal derfor godtgøre, om de udvalgte 31 skoler i stedet for lever op til krav som om de var en tilfældig stikprøve fra starten, dvs. krav om at være *proportionale* i forhold til landsdækkende oplysninger for en række baggrundsvariable. På denne måde sikrer en statistisk analyse, at den benyttede stikprøve ikke tegner et skævt billede af de endelige normer for prøvepræstationerne, set i forhold til en række baggrundsvariable, som man ved, har systematisk indflydelse på prøveresultaterne

Bortset fra den måde, som skolerne er udvalgt på er den samplingsteknik, som har været anvendt den samme, som anvendes ved de internationale undersøgelser OECD’s PISA og IEA’s TIMSS og PIRLS, som medtager i alt ca. 4000 elever pr studie. I disse studier sikrer princippet om simpel tilfældig udvælgelse automatisk repræsentativitet af de udvalgte elever.

 En såkaldt totrins-sampling med successiv udvælgelse først af skoler og derefter af elever inden for skolen medfører normalt (internationale studier), at eleverne skal tildeles vægte, der fx sikrer, at vejede gennemsnit ikke er ’skæve’ og samtidig er mere præcise, end hvis de gennemføres som uvejede gennemsnit. Tekniske overvejelser omkring variationen i skolestørrelse og klassestørrelse har ført til, at denne type vægtning er udeladt af de statistiske analyser.

De skoler, der deltog, er vist i tabel 1sammen med oplysninger om skolen har bidraget med data til sætningslæseprøverne SLP1 og 2, fra både forår og efterår.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Skolenavn | Placering | DeltagerSLP1 forår | DeltagerSLP2 forår | DeltagerSLP1 efterår | DeltagerSLP2 efterår |
| Brovst Skole | 9460 Brovst | X | x | x | X |
| Egedalsskolen | 2980 Kokkedal | X | x |   |   |
| Gildbroskolen | 2635 Ishøj |   |   | x | X |
| Gudumholm Skole  | 9280 Storvorde  | X | x | x | X |
| Gødvadskolen | 8600 Silkeborg | x | x | x | X |
| Hesselager skole | 5874 Hesselager | x | x | x | x |
| Hvinningdalskolen | 8600 Silkeborg | x | x | x | x |
| Højstrupskolen | 5200 Odense V | x | x | x | x |
| Ishøj Skole | 2635 Ishøj | x | x | x | x |
| Kvaglundskolen | 6705 Esbjerg Ø | x | x |   |   |
| Lille Værløse Skole | 3500 Værløse | x | x | x | x |
| Lyngbjerggaardskolen | 9230 Svenstrup | x | x | x | x |
| Løgumkloster Kommuneskole | 6240 Løgumkloster | x | x | x | x |
| Løjtegårdsskolen | 2770 Kastrup | x | x |   |   |
| Marienlystskolen | 3600 Frederikssund | x | x | x | x |
| Nordstrandskolen | 2791 Dragør | x | x | x | x |
| Ravnsholtskolen | 3450 Allerød | x | x | x | x |
| Rise Skole  | 5970 Ærøskøbing | x | x |   |   |
| Sct. Nicolai Skole | 4600 Køge |   |   |   |   |
| Stengårdsskolen | 3650 Ølstykke | x | x | x | x |
| Strandgårdsskolen | 2635 Ishøj | x | x | x | x |
| Sæbygårdskolen | 9300 Sæby | x | x |   |   |
| Søholmskolen | 4174 Jystrup |   |   |   |   |
| Søndervangskolen | 8260 Viby J | x | x | x | x |
| Tirsdalens Skole | 8900 Randers | x | x | x | x |
| Tranbjerg Skole | 8310 Tranbjerg J |   |   | x | x |
| Tranebjerg skole | 8305 Samsø | x | x | x | x |
| Tårnborg Skole  | 4220 Korsør | x | x | x | x |
| Vinderslev Skole | 8620 Kjellerup | x | x | x | x |
| Ådalskolen | 6710 Esbjerg V |   |   | x | x |

**Tabel 1** Opgørelse af 31 skoler, som er udvalgt til at deltage samt deres status mht. til at levere data forår og efterår.

Det fremgår af tabellens skraverede felter, at der er skoler, som ved et par lejligheder ikke har indleveret data. Selv om en udregnet svarprocent på elevniveau generelt viser en relativ høj værdi, er der på grund af det ikke-tilfældige udgangspunkt med de 80 skoler alligevel grund til at undersøge, om de deltagende 31 skoler bidrager med data, der er ’skæve’ i den betydning, at de eventuelt trækker gennemsnit og normer i en ’skæv’ retning i forhold til en stikprøve, der kunne vælges som en tilfældig stikprøve blandt alle Danmarks ca. 1500 skoler. Denne type analyse kan i sagens natur ikke gennemføres ud fra beregnede elevfærdigheder. Men den kan gennemføres ved hjælp af en række eksterne karakteristika, som ligger som ’baggrundsoplysninger’ for de indsamlede data.

I korthed skal man undersøge, om visse eksterne *karakteristika* er repræsenteret i den faktisk opnåede stikprøve med skoler med samme hyppighed som de er repræsenteret i den totale population, hvorfra stikprøven i princippet stammer fra. Altså i forhold til forholdene inden for *samtlige* ca. 1500 skole i Danmark. Fx skal det undersøges om den faktiske stikprøve er ’skæv’ mht. fordelingen på køn af lærere.

De *karakteristika*, som er til rådighed og kan anvendes er

1. Kønsfordeling af lærere
2. Aldersfordeling af lærere
3. Skolestørrelse
4. Hyppighed af tosprogede
5. Geografisk spredning af skolerne

Analyserne gennemføres for de skoler, der har leveret data forår.

**Geografisk spredning**

Det sidste punkt, den geografiske spredning handler ikke om at eftervise, at stikprøvens skoler ligger helt tilfældigt fordelt i Danmark, fordi virkeligheden ikke ser sådan ud. En mulighed er at opgøre antallet af skoler i stikprøven i forhold til de fem regioner, som Danmark er delt op i, og derefter sammenligne med fordelingen af elever i de udvalgte skoler over regionerne. Det er gjort i tabel 2

|  |  |
| --- | --- |
|  Region | Procent elever i 31 deltagende skoler af regionens elevantal |
| Hovedstaden | 2.97% |
| Midtjylland | 1.98% |
| Nordjylland | 2.31% |
| Sjælland | 1.27% |
| Syddanmark | 1.51% |

**Tabel 2** Andelen (procent) elever i udvalgte 31 skoler af regionens elevantal, fordelt på samtlige Danmarks 1591 folkeskoler

De fem procentværdier sammenlignes ud fra et sædvanligt χ 2 tests for kontingenstabeller. Testet afviser, at eleverne fra de deltagende skoler er konstant fra region til region. Der er signifikant flere elever fra Hovedstaden og Nordjylland sammenlignet med de øvrige regioner. Afvigelserne skønnes imidlertid at være så numerisk små, at de *ikke* at har betydning for udarbejdelsen af normer på baggrund af de 31 skoler.

**Kønsfordeling af lærere**

Iflg. Officielle statistikker fra Danmarks Statistik (DST) udgør andelen af mandlige lærere i den danske folkeskole i øjeblikket ca 32%. Fra skoleåret 2002 til 2006 falder andelen af mænd på folkeskoler fra 34.3 til 32.7 procent og DST’s sidste tal er 31.5%.

I stikprøven er andelen af mandlige lærere 30.0%. Denne andel afviger ikke signifikant (binomial test) fra tallet 31.5% (p>0.05) og konklusionen er derfor at forårets kønsfordeling for lærerne svarer til virkeligheden i Danmarks folkeskole.

**Aldersfordeling af lærere**

Den gennemsnitlige alder for lærere i folkeskolen har ligget stabilt på 44.4 år over årene 2003-2009. I stikprøven er den gennemsnitlige læreralder 44.3. Dette tal afviger ikke signifikant fra gennemsnitstallet på landsplan (ANOVA[[1]](#footnote-1), p>0.05) og stikprøvens aldersgennemsnit svarer derfor til det totale gennemsnit på landsplan.

Et lidt mere detaljeret billede fås ved at se på selve fordelingen (grupperet) af læreralder,

tabel 3 viser disse detaljer.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data kilde | 18-24 år | 25-34 år | 35-44 år | 45-54 år | Over 54 år | I alt procent |
| Stikprøve | 0.0% |  0.0% | 46.2% | 53.8% |  0.0% | 100.0% |
| Danmark 2007 | 0.5% | 25.1% | 23.9% | 25.9% | 24.6% | 100.0% |

**Tabel 3** fordelingen af læreralder (grupperet) i stikprøven og for samtlige lærere i Danmarks 1591 folkeskoler i 2007.

Et χ 2 tests afslører, at der er signifikant flere lærere i aldersgruppen 35-44 og 45-54 år i stikprøven sammenlignet med forholdende for hele landet. De yngste og de ældste lærere er derfor ikke repræsenteret i stikprøven, selv om den gennemsnitlige alder svarer til landet som helhed.

**Skolestørrelse**

Den gennemsnitlige skolestørrelse i stikprøven er 386 elever. Ifølge Danmarks Statistik er den gennemsnitlige folkeskolestørrelse 377 elever. Stikprøvens 386 elever afviger ikke signifikant fra tallet som gælder de samlede 1591 folkeskoler i Danmark (approximativ ANOVA, p>0.05)

**Hyppighed af tosprogede**

Den gennemsnitlige hyppighed af tosprogede elever i stikprøven er 16.86%. Til sammenligning oplyser Danmarks Statistik det seneste gennemsnitstal for tosprogede i den samlede folkeskole på 10.1% . Til trods for den relativt store forskel, afviger 16.86% ikke signifikant fra værdien 10.1% gældende for hele landet (logit transformation af procentværdierne , derefter ANOVA)

Lidt flere detaljer kan man få ved at se på fordelingen af skoler i grupperinger af hyppigheden af tosprogede. Det er gjort i tabel 4.

|  |  |  |
| --- | --- | --- |
| Hyppighed af tosprogede | Pct skoler i Danmark | Pct skoler i stikprøve |
| 0.0-0.0 | 8.4 |  7.7 |
| 0.1- 9.9 | 67.3 | 53.9 |
| 10.0-24.9 | 15.0 | 11.6 |
| 25.0-34.9 | 4.1 | 11.6 |
| 35.0-49.9 | 2.6 | 7.7 |
| 50.0-74.9 | 1.7 | 3.9 |
| 75.0-100.0 | 1.4 | 3.9 |

**Tabel 4** fordelingen af skoler i grupperede værdier af andel tosprogede elever i skolen

Et χ 2 tests på det mere detaljerede niveau ud fra tabel 4’s oplysninger bekræfter testet for den gennemsnitlige værdi: Der er ikke signifikant forskel på andelen af tosprogede i stikprøven sammenlignet med forholdene blandt Danmarks 1591 folkeskoler.

*Ved hjælp af de gennemførte frafaldsanalyser kan det således konkluderes, at forårets sample har ’bevaret’ den principielle repræsentativitet, som den blev skabt ved via den simple tilfældige stikprøveindsamling*

**Sammenligning af forår og efterår mht. til skiftende deltagelse af skoler**

Som det fremgår af tabel 1 er der skoler, som kun delvist er med ved den samlede afprøvning forår og efterår. Fx ses det, at der er 5 skoler, som ikke er med ved forårets dataindsamling, mens der er 7 skoler, som ikke er med i efterårets samplingsrunde. To af skolerne deltog hverken forår eller efterår.

For at vurdere om det har haft betydning for repræsentativiteten af efterårets sampling undersøges det, om de skoler, der har de facto har leveret data til efteråret (ikke-skraverede i søjle 5 og 6 i tabel 1) adskiller sig fra (a) de skoler, som *burde* levere data til efteråret (dvs alle skoler) og (b) de skoler, som de facto leverede data fra foråret (dvs skoler i søjle 3 og 4 i tabel 1). Kriterierne for ’at adskille sig’ er en gentagelse fra frafaldsanalysernes karakteristika:

1. Kønsfordeling af lærere
2. Aldersfordeling af lærere
3. Skolestørrelse
4. Hyppighed af tosprogede

Den geografiske repræsentativitet kan ikke undersøges meningsfuldt med de få skoler der skifter ’status’ fra forår til efterår.

De nævnte kriterievariable a - d er alle kvantitative variable, men uden indlysende statistisk fordeling. Derfor undersøges problemet om ’at adskille sig’ ved ikke-parametriske tests (Wilcoxon).

*Resultatet af disse tests er, at der ikke optræder signifikante forskelle overhovedet. Det betyder, at ser man på repræsentativiteten gennem briller der defineres ud fra niveauerne af de nævnte karakteristika a – adskiller de facto stikprøven i efteråret sig ikke fra hverken den stikprøve af skoler, som burde have leveret data om efteråret eller de skoler, som de facto har leveret data ved forårets sampling.*

**Sætningslæseprøverne 1 og 2**

Opgaverne er placeret i to prøvehefter og repræsenterer læsning af sætninger af stigende sværhedsgrad. Prøvetid er for begge hæfter 8 minutter. Tilsammen søger de to prøver at dække eleven og klassens udvikling i sætningslæsning hvert år fra begyndelsen af 1. klasse til sidst i 5. klasse.

Sætningslæseprøvernes opgavetype er alle af multiple-choice typen (MC) med en række sætninger knyttet til hvert billede. Sværhedsgraden går systematisk fra korte sætninger med korte ord til længere og mere sammensatte sætninger med gradvist sværere ord. Dermed er hensigten at prøveresultaterne kan beskrive hvert enkelt niveau i sætningslæsningens udvikling.

Sætningslæseprøve1 består af 26 opgaver med 4 svarmuligheder per opgave og sætningslæseprøve2

består af 27 opgaver, også her med 4 svarmuligheder per opgave. Tiden til løsningen af læseopgaverne er løbende registreret via farveskift efter to minutter, fire minutter, seks minutter osv.

Data er indsamlet på de skoler, som er nævnt i tabel 5, hvor det ses, at det langt fra har været muligt at få dækket samtlige prøver med samtlige klassetrin på skolen

| **Sætningslæseprøve SLP** | **klassetrin** |
| --- | --- |
| **1** | **2** | **3** | **4** | **5** |  |
| **N= antal elever** |
| **skole** | **SLP** | 23 | 18 | . | . | . |  |
| **Amager Fælled Skole** | **1** |
| **2** | . | 18 | 24 | 16 | 14 |  |
| **Bakkegårdsskolen** | **1** | . | 21 | . | . | . |  |
| **2** | . | 21 | . | . | . |  |
| **Brovst skole** | **1** | 20 | 24 | . | . | . |  |
| **2** | . | 24 | 18 | 18 | 13 |  |
| **Egedalsskolen** | **1** | 17 | 22 | . | . | . |  |
| **2** | . | 21 | 18 | 20 | 17 |  |
| **Gildbroskolen** | **1** | 16 | 19 | . | . | . |  |
| **2** | . | 18 | 16 | 21 | 20 |  |
| **Gudumholm Skole** | **1** | 18 | 24 | . | . | . |  |
| **2** | . | 24 | 12 | 20 | 26 |  |
| **Gødvad skole** | **1** | 23 | 23 | . | . | . |  |
| **2** | . | 24 | 21 | 22 | 20 |  |
| **Hesselager Skole** | **1** | 14 | 14 | . | . | . |  |
| **2** | . | 14 | 16 | 17 | 13 |  |
| **Hvinningdal skole** | **1** | 23 | 24 | . | . | . |  |
| **2** | . | . | 24 | 25 | 23 |  |
| **Højstrup skole** | **1** | 15 | 12 | . | . | . |  |
| **2** | . | 12 | 17 | 16 | 18 |  |
| **Ikke oplyst** | **1** | 15 | 17 | . | . | . |  |
| **2** | . | 31 | . | . | . |  |
| **Ishøj Skole** | **1** | 19 | 21 | . | . | . |  |
| **2** | . | 23 | 22 | 19 | 22 |  |
| **Kvaglundskolen** | **1** | 24 | 15 | . | . | . |  |
| **2** | . | 14 | 15 | 20 | 15 |  |
| **Lille Værløse skole** | **1** | 17 | 19 | . | . | . |  |
| **2** | . | 19 | 21 | 22 | 20 |  |
| **Løgumkloster kommuneskole** | **1** | 24 | 19 | . | . | . |  |
| **Løjtegårdskolen** | **1** | 22 | 20 | . | . | . |  |
| **2** | . | 20 | 20 | 24 | 24 |  |
| **Marienlyst Skole** | **2** | . | . | . | . | 17 |  |
| **Nordstrandskolen** | **1** | 24 | 20 | . | . | . |  |
| **2** | . | 20 | 11 | 14 | 24 |  |
| **Ravnsholt Skole** | **1** | 14 | 19 | . | . | . |  |
| **2** | . | 18 | 32 | 20 | 28 |  |
| **Rise Skole** | **1** | 12 | 9 | . | . | . |  |
| **2** | . | 13 | 18 | 12 | 18 |  |
| **Stengården skole** | **1** | 20 | 17 | . | . | . |  |
| **2** | . | 17 | 38 | 21 | 17 |  |
| **Strandgårdskolen** | **1** | 19 | 21 | . | . | . |  |
| **2** | . | 18 | 18 | 15 | 14 |  |
| **Sæbegårdskolen** | **1** | 24 | 19 | . | . | . |  |
| **Søndervangskolen** | **1** | 16 | 13 | . | . | . |  |
| **2** | . | 32 | . | 23 | 15 |  |
| **Tirsdalens Skole** | **1** | 20 | 18 | . | . | . |  |
| **2** | . | 18 | 23 | 15 | 17 |  |
| **Tranbjerg Skole** | **1** | 33 | 35 | . | . | . |  |
| **2** | . | 21 | . | 19 | 14 |  |
| **Tårnborg Skole** | **1** | 9 | 16 | . | . | . |  |
| **2** | . | 17 | 16 | 21 | 14 |  |
| **Vinderslev Skole** | **1** | 23 | 26 | . | . | . |  |
| **2** | . | 20 | 18 | 18 | 18 |  |
| **Ådalskolen** | **1** | 28 | 19 | . | . | . |  |
| **2** | . | 19 | 18 | 16 | 20 |  |

**Tabel 5** Antallet af elever, som har deltaget på skolerne i de to sætningslæseprøver 1= efterår og 2= forår

Beregning af et mål for løsningssikkerhed finder sted ud fra to forskellige grundtanker (se vejledningen for detaljer) : (1) Ud fra procent rigtige *pct\_ri*, defineret som antallet af rigtige svar i forhold til antallet af opgave, som eleven har nået at se på (antal passerede) og (2) ud fra at

sammenholde antallet af korrekt besvarede med den tid, der er benyttet *hj\_ri*. Der er tilsvarende to måder at beregne den medgåede løsnings tid (se vejledningen) på, som medfører to forskellige mål for *løsningshastigheden*: (1) *has\_tid* og *hj\_tid*, som tager afsæt i de to forskellige måder at registrere løsningstiden på.

|  | **klassetrin** |
| --- | --- |
| **1** | **2** | **3** | **4** | **5** |
| **hj\_ri** | **hj\_ri** | **hj\_ri** | **hj\_ri** | **hj\_ri** |
| **gsnit** | **gsnit** | **gsnit** | **gsnit** | **gsnit** |
| **t\_hj\_tid** | **SLP** | 0.90 | 0.93 | . | . | . |
| **1** | **1** |
| **2** | . | 0.90 | 0.92 | 0.89 | 0.93 |
| **2** | **1** | 0.89 | 0.95 | . | . | . |
| **2** | . | 0.94 | 0.95 | 0.96 | 0.97 |
| **3** | **1** | 0.81 | 0.96 | . | . | . |
| **2** | . | 0.86 | 0.97 | 0.95 | 0.96 |

|  | **klassetrin** |
| --- | --- |
| **1** | **2** | **3** | **4** | **5** |
| **pct\_ri** | **pct\_ri** | **pct\_ri** | **pct\_ri** | **pct\_ri** |
| **gsnit** | **gsnit** | **gsnit** | **gsnit** | **gsnit** |
| **t\_has\_tid** | **SLP** | 0.75 | 0.95 | . | . | . |
| **1** | **1** |
| **2** | . | 0.82 | 0.95 | 0.92 | 0.94 |
| **2** | **1** | 0.89 | 0.94 | . | . | . |
| **2** | . | 0.93 | 0.94 | 0.95 | 0.96 |
| **3** | **1** | 0.85 | 0.90 | . | . | . |
| **2** | . | 0.89 | 0.92 | 0.91 | 0.95 |

**Tabel 6** *Løsningssikkerhed*  (procent rigtige) beregnet som *pct\_ri* og *hj\_ri* sammenholdt med *løsningshastigheden* beregnet som *has\_tid* og *hj\_tid* (tidsforbrug, se vejledning)

 Informationen i tabel 6 kan gengives grafisk uden den anvendte gruppering af løsningshastigheden. Figur 1 gengiver relationen for alle klassetrin, set under ét og for begge SLP – prøver på én gang. Den karakteristiske krumning, som begge figurer viser, er grundlaget for en klassisk vurdering af eleverne i et todimensionalt kategorisystem (se nærmere i vejledningen).





**Figur 1** *Løsningssikkerhed* sammenhold med løsningshastighed under brug af de to forskellige beregningsgrundlag (1) og (2). Alle klassetrin er repræsenteret og begge sætningslæseprøver er gengivet på én gang med forskellige farver og type: SLP1 og 2, hver for sig, angivet med type 1=Blå og type 2=brun

A**nalyser af homogenitet af opgaver ved hjælp af Rasch modeller**

Opgaverne i hver af sætningslæseprøverne er konstrueret med henblik på at måle én bestemt sætnings-læsefærdighed hos eleverne. For at dette kan lade sig gøre stiller der, set fra en psykometrisk vinkel et krav om éndimensionalitet, dvs et krav om, at hver opgave bidrager på lige fod til en samlet elevscore, som derefter er målet for elevfærdigheden. Grundlaget for analyse af, om sådanne krav konkret opfyldes af opgaverne i SLP1 og SLP2 er en omfattende kontrol af, om den såkaldte Rasch Model passer på de indsamlede data.

I det følgende beskrives kort, hvad denne analyse indeholder. Det empiriske udgangspunkt er data fra de 31 skolers elever. En skabelon for sådanne data er gengivet i tabel 7.

|  |  |  |
| --- | --- | --- |
|  | Opgaver |  |
| Elever | Opg. 1 | Opg. 2 | Opg. 3 | Opg. I | . | Opg. k | Elevscore(rv) |
| 1 | 1 | 0 | 1 | 1 | 1 | 0 | a1. |
| 2 | 0 | 1 | 1 | 0 | 0 | 1 | a2. |
| 3 | 1 | 1 | 0 | 1 | 1 | 0 | a3. |
| . |  |  |  |  | . |  |  |
| V | 1 | 0 | 1 | avi | . | 1 | av. |
|  |  |  |  |  |  |  |  |
| N | 1 | 1 | 0 | aNi | . | 0 | aN. |
| Item score (si) | a.1 |  |  | a.i | . | a.k | a.. |

**Tabel 7:** En generel ramme for registrering af elevers svar avi på SLP1 og SLP2-lignende prøver.

Tabel 7 sammenfatter N elevers svar på **k** opgaver, som prøven tænkes at bestå af. Elev nr. 1 har svaret rigtigt på opgave nr. 1 (1), forkert på opgave nr. 2 (0) osv., idet man benytter *kodningen* 1 = ’rigtig besvarelse’, 0 = ’forkert besvarelse’. I alt har elev nr. 1 haft et antal rigtige besvarelser: a1., som er *elevscoren*. Man kan bemærke, at med en scoring 1/0 for rigtig/forkert besvarelse, sikrer man sig en ”nem” måde at beregne scoren på, nemlig som *summen* af de kodede besvarelser. Det giver nogle forenklinger, når man regner på tabellen rent matematisk. Men kodningen kunne lige så godt have været: 1 = ’rigtigt’, 2 = ’forkert’. Under alle omstændigheder er der tale om, at selve besvarelsen ’rigtig’ eller ’forkert’ i virkeligheden er et *kategoriseret udfald* uden numeriske værdier knyttet til sig.

I Tabel 1 har elev nr. 1 svarmønstret (1,0,1,1,1,0) – til forskel fra elev nr. 3, som har svarmønstret (1,1,0,1,1,0). Begge elever har 4 rigtige i alt – elevscoren av. er lig med 4 – og man kan betragte de to besvarelser som ’forskellige’ eller ’ens’ afhængig af om man noterer sig svarene på de enkelte opgaver eller om man alene ser på antallet af rigtige svar, elevscoren.

På trods af, at opgavebesvarelserne har et diagnostisk formål, der synes at placere den enkelte elev i en kategori helt for sig selv, defineret ved hjælp af mønstret af 1’ere og 0’ler hen over de **k** opgaver, er der overordnet tale om, at man ønsker at sammenligne eleverne ud fra en simpel *rangordning* ved hjælp af elev-scorerne.

Mulighederne for at kunne foretage sådanne simple sammenligninger ud fra antallet af korrekte svar (elev scoren) blev af G. Rasch (Rasch,1960) ’oversat’ til at de data, som er indeholdt i tabel 1 skal kunne beskrives ved hjælp af den såkaldte Rasch Model:

Følgende (logistiske) én-partameter model er den simpleste og mest krævende model af Rasch modellerne: θ1,..., θ2, måler opgavesværhederne og σ1,..., σn, måler elevernes færdighed. Med disse to sæt teoretiske mål (sværhed = θi og færdighed= σv) er sandsynligheden for et korrekt svar til item nr. i fra elev nr. v (dvs. avi=1) følgende (1):

 (1)

Denne model er speciel blandt mængden af IRT (Item Response Theory) modeller. Det skyldes bla. At Rasch (Rasch,1968) viste, at tre følgende udsagn er ækvivalente: (i): elevscorerne (og item scorerne) udtømmer *al* viden om ’sværhed’ og ’dygtighed’ (sufficiens), (ii): Det er muligt at sammenligne elevfærdigheder med en hvilken som helst subgruppe af items (skal føre til samme resultat) og (iii) Rasch modellen er en gyldig statistisk beskrivelse af data i tabel 1.

Den første egenskab kan kaldes en ‘validering’ af den praktiske anvendelse af elevscorer. Den anden egenskab kaldes ’specifik objektivitet’ og er ekstrem anvendelig ved testsituationer, hvor ikke alle elever får de same opgaver (natioionale tests, PISA og IEA’s TIMSS og PIRLS undersøgelser)

Rent praktisk kontrolleres Rasch modellens gyldighed (iii), også i dette tilfælde med OLP1 og OLP2 ved at studere de såkaldt item karakteristiske kurver ICC(σv):

 (2)

Disse kurvers S-form skal se ud på en bestemt made for at modellen har gyldighed.

Undertiden benyttes den såkaldt to-parameter (logistiske) Rasch model (3), hvis ikke den simple én-parameter model viser sig at have gyldighed (Lord and Novick ,1968):

**** (3)

I model (3) indgår item diskriminations parametre δ1,...., δn ved siden af itemparametrene, de markerer hældninger på de item karakterisiske kurver ICC's og skal estimeres som et ‘ekstra’ sæt parametre, sammen med θ - og σ parametre. Jo stejlere en ICC kurve er jo bedre diskrimineres mellem naboværdier af elevfærdigheder. Modellen (3) reduceres til den simple model (1), hvis alle item diskriminationer er lig med 1: δi=1 .

**Beregning af elevfærdigheder**

Elevfærdighederne σ1,..., σn skal beregnes og det er i princippet dem, der indgår i normer fra den anvendte test (PISA scores og IEA scores fra TIMSS og PIRLS)

Det er nødvendigt at have estimeret item parametrene først, før man estimerer elvernes færdighedsniveauer σ1,..., σn . Det foregår ved at indsætte de estimerede itemparametre θ1,..., θk i følgende (Maximum Likelihood) ligning (4) :

 (4)

 (8)

Venstre siden r angiver elevscoren. Der er ikke løsninger til ligningen for r=0 og r=k, altså for ‘ingen korrekte besvarelser’ eller for ‘alt rigtigt’

Vurdering af specifikke elevsvar kan desuden finde sted ved at se på følgende sandsynlighed for at får en given svarsekvens (avi(0),...,avK(0)) givet at eleven har av.(0) rigtige i alt:

 (9)

Hvis sandsynligheden er mindre en 0.05 bør man se efter, hvad der kunne være årsagen. Det kunne dække over, at der er såkaldt ’lokal’ afhængighed i svarene til de enkelte items. Altså at et svar på ét item kan afhænge af, om der er svaret rigtigt eller forkert på det foregående item.

Den præcision, hvormed elevernes færdigheder estimeres, afhænger af antallet af opgaver, som eleven besvarer. En tommelfingerregel, som bla. følges ved afviklingen af De Nationale Adaptive IT-baserede Tests er, at eleverne mindst skal besvare et antal opgaver, således at *standard error of measurement* (SEM) er mindre end 0.3, Denne størrelse regnes simpelt ud fra følgende relationer



 Fortolkningen af den beskrevne målefejl (errror of measurement) refererer til at eleven ’gentager’ besvarelser over flere items.

 Umiddelbart skulle man tro, at en beregning af en tilsvarende ’fejl’ på beregningen af elevens læsehastighed ville være umulig, fordi der her alene er et samlet antal besvarede og en anvendt tid til rådighed. Men for hver elev findes der i disse data en *hel række* af angivelser af hvor langt eleven er nået efter 1,2,3,4 ..osv minutter. Disse oplysninger er, som antydet ovenfor omregnet til mål for elevens læsehastighed (antal besvarede opgaver pr 2 minutter). Nogle elever har nogle få målinger af læsehastighed, andre elever har op til 8 successive målinger, afhængig af, hvor langt eleverne er nået inden for de totalt fastsatte tidsrammer. Det nødvendige element af ’gentagelse’ for beregning af en usikkerhed på læsehastigheden ligger derfor i disse successive oplysninger. Analyser af de successivt beregnede læsehastigheder viser, at læsehastigheden systematisk går ned med antallet af målinger, dvs med antallet af besvarede opgaver. Denne ’trend’ blev lineært tilpasset via en simpel regression, og målingerne blev korigeret for denne systematik, før standardafvigelser på de korrigerede læasehastighedsmål blev lagt til grund for en vurdering af usikkerhed på elevens læsehastighed.

**Andre aspekter af valideringen af sætningslæseprøverne og konstruktion af normer**

Den samlede vurdering af en elevs præstationer skal, som omtalt, både referere til et mål for *læsesikkerheden* og et mål for *læsehastigheden*. I figur 2 er grundlaget for denne vurdering illustreret (jf figur 1, med omvendte akser)



**Figur 2** Relation mellem elevers læsehastighed (Y-akse) og læsesikkerhed (X-akse).

Figur 3 viser en velkendt, karakteristisk ’banan’ form, der placerer majoriteten af elever langs med X-aksen for neden og lodtret op ad Y-aksen til højre, dvs. som ’sikre’ og ’hurtige’ læsere eller som ’sikre’ men ’langsomme’ læsere – og endelig, som ’usikre’ og ’langsomme’ læsere

I figur 2 er desuden antydet nogle rektangler svarende til 95% konfidensgrænser lagt pr. elev efter de udregningsprincipper, som der er redegjort for ovenfor. Standardafvigelser på læsehastigheden markeres lodret (faktor 2 svarende til approksimativ afgrænsning af 95% i en normalfordeling med ±2 standardafvigelser) mens standardafvigelsen på læsesikkerheden er markeret vandret ( også med faktor 2). Udregningerne viser at begge standardafvigelser er af størrelsesordenen ±5.

Det følger almindelige statistiske principper, at hvis man vil regne standardafvigelsen ud på et ’klasseresultat’ fx over ca. 20 elever bliver usikkerheden udtrykt ved et konfidensinterval med ’bredden’ og ’højden’ ± 5/4.47, hvor 4.47 er kvadratroden af 20.

**Konklusion**

De statistiske analyser har vist, at sætningslæseprøverne i deres nuværende form lever op til gængse psykometriske krav vedrørende reliabilitet og validitet, hvad angår kravet om, at enkeltopgaver udgør et homogent grundlag til at bygge skalaer til måling af færdigheder med (Rasch homogenitet). Sætningslæseprøverne producerer mål for elevernes læsesikkerhed og læsehastighed, som er tilfredsstillende set i forhold til gængse krav om lille usikkerhed på de opnåede mål .

**Referencer**

Allerup, P. (1994) Rasch Measurement, Theory of. *The International Encyclopedia of Education*; second edition, Pergamon Press (1994).

Allerup,P, Jansen,M, Weng,P *Evaluering i skolen- baggrund, praksis og teori*, Dafolo 2011

Andersen E.B.(1973) Conditional Inference and Models for Measuring. *Mentalhygiejnisk Forlag*. Copenhagen.

Elley, W.B.(1992). How in the World do Students read.

 *The International Association for the Evaluation of Educational Achievement*, The Hague.

Lord F.M.(1980) Applications of Item Response Theory to Practical Testing Problems. *Lawrence Erlbaum Associates*, New Jersey.

Lord F.M. and Novick M. R.(1968). Statistical Theories of Mental Test Scores. *Addison Wesley*, Massachusetts.

Lundberg I. and Linnakyla P.(1992) Teaching Reading around the World. *The International Association for the Evaluation of Educational Achievement*, The Hague.

Postlethwaite T.N. and Ross K.(1992) Effective Schools in Reading, implications for educational

 planners. *The International Association for the Evaluation of Educational*

 *Achievement*, The Hague

1. ANOVA anvendes approximativt, selv om aldersfordelingen har en tendens til at være to-puklet [↑](#footnote-ref-1)