

Karakterdannelsen

Problemidentifikation	1
Empirisk analyse.	3
Analysemetode	3
Analyse af gennemsnitskarakteren	4
Analyse af karakterspredningen	8
Sammenfattende bemærkninger	9
Kommentar til karakterkommissionen	13

Problemidentifikation

Er karaktergivning "tacit knowledge" – en form for medfødt gave? Når man ser sig om i det pædagogiske landskab kan man nemt få den tanke. Her er det uhyre vanskeligt at finde denne opgave behandlet – hverken i faglige eller i debatmæssige sammenhænge. Eller måske er emnet bare ikke rigtigt komme il faut i konsensusmødet? Bedømt ud fra love, bekendtgørelser, cirkulærer mv. er der ikke tvivl om, at der fra den styrende statsmagts side tillægges evaluering en endog særdeles væsentlig betydning. Udmøntningen heraf i praktiske redskaber, anvendelige vejledninger eller brugbare instruktioner er derimod meget mere sjældne ligesom en vidensspredning gennem erfaringsudveksling eller en "tavs" mesterlære stort set er non-existent.

I det større billede er karaktergivning det synlige element i den eksterne evaluering, og på det over-ordnede plan skal der tages stilling til såvel evalueringens funktion (hvorfor) som til dens metode (hvordan). Medens funktionen er det centrale i de ministerielle tiltag, er metodeaspekterne anderledes stedmoderligt behandlet – og det gælder i alle styrende instanser. Metodemæssigt står man overfor to hovedproblemer, nemlig

- a. evalueringens brugbarhed eller validitet
- b. evalueringens pålidelighed eller reliabilitet

Der er naturligvis stor samhörighed mellem de to problemer, men dog således at effekten hovedsagelig går fra b mod a – en evalueringens brugbarhed er langt hen ad vejen bestemt af dens pålidelighed - bramfrit udtrykt ved garbage in, garbage out. Når evalueringens pålidelighed er lav er brugbarheden tilsvarende beskeden eller måske endda helt uanvendelig.

En evalueringens manglende reliabilitet – hvad enten den betegnes som usikkerhed, inkonsistens, nonkohærens eller noget helt fjerde – kan fremkomme af mange årsager. Følgende tre forhold vil dog omfatte de væsentligste.

- proceduremæssige svagheder. Mangel på præcise og velbeskrevne mål og kriterier for bedømmelsen vil medføre at bedømmelsen kommer til at bygge på tilfældige skøn frem for præcise krav. Hvis der blandt bedømmerne hersker vidt forskellige opfattelser af vurderingskriterierne – herunder en vægtning af samme – er det alene tilstrækkeligt til at undergrave evalueringens pålidelighed.

- person-specifikke træk. Den måde man vurderer på er i betydelig grad afhængig af personlige træk hos bedømmeren. Nogle mennesker er i almindelighed strenge bedømmere og stiller store krav medens andre er mere milde og tolerante. Nogle mennesker har en personlighed, der gør at de holder sig til midten og det moderate, medens andre gerne bevæger sig ud til siderne. Personlig erfaring og uddannelse er blandt de erhvervede egenskaber, der kan påvirke bedømmeren. Virkningen af sådanne person-specifikke træk vil naturligvis også øve indflydelse på evalueringens pålidelighed.
- vilkårligt forekommende forhold. Under disse betegnelser gemmer sig et stort antal vidt forskellige forhold. Stærkt varierende karaktermiljø, manglende eller underudviklede redskaber, ydre forstyrrelser, træthed, latente konfliktforhold, forventninger, misforståelser osv., kan forekomme i et omfang, der ofte er umulig at kontrollere og tage højde for i en evalueringssituation. Et uudtrykt men stærkt ønske fra skolens ledelse om størst mulig gennemførelsesprocent hører muligvis ikke under vilkårligt forekommende forhold, men virkningen er den samme for bedømmeren og vil som sådan også påvirke vurderingens pålidelighed.

Tilbage – og alene – står den enkelte lærer med ikke ret meget andet end sine egne formodninger, personlige erfaringer og subjektive tolkninger af fagenes generelle mål som evalueringgrundlag. Det er utvivlsomt helt korrekt når såvel lærere som skoler hævder at der ikke eksisterer et karaktergivningsproblem på deres skole. Men denne påstand skal ses på baggrund af, at der er en udbredt mangel på faktisk viden om og erkendelse af hvordan karakterdannelsen overhovedet finder sted. Ideelt set – og også det der forventes af de karakterbrugende interessenter – er karaktergivningen *lærer- og situationsuafhængig samt absolut*, og det er i denne kontekst, der kan være problemer, som i større eller mindre omfang reducerer værdien af karaktergivningen. Skolernes forventning til og lærernes egenopfattelse af, at de afgivne karakterer er givet efter bedste overbevisning, er imidlertid ikke tilstrækkelig sikkerhed for unbiased karakterer. Selvom den enkelte lærer er helt entydig i sin karaktervurdering, kan der være uerkendte og nærmest usynlige, men gennemgående skævheder i karakterdannelsen og som kan medføre større eller mindre skævvridninger af den samlede karaktersum.

Skolesystemet som sådant indeholder ingen feed-back, ingen korrektivanvisninger eller vurderinger – hvad enten den er positiv eller negativ - intet tilbud om udvikling, ingen anvisninger på eller forskrifter for hvordan, ingen hjælpemidler til brug for lærernes karaktergivning eller egenvurdering heraf – kort sagt er det en helt privat sag for den enkelte lærer – lige indtil en klasses årskarakter kommer under 6 eller der indløber en klage – så bliver skolen nødt til at rykke ud som en anden brandmand. Den proaktive indsats er dog at foretrække, men emnets ømtålelighed, skolerne manglende viden og erfaring på området og en nem form for ansvarsfraskrivelse taler alle for, at det er meget nemmere og mere bekvemt at nedprioritere problemstillingen, indtil der så at sige kommer et uafviseligt udbrud fra undergrunden.

For den enkelte lærer bliver en vis og måske som regel ubevidst "taktisk karaktergivning" derfor en nærliggende – og måske også eneste – mulighed. Det kan gå på forskellige linier – lige fra 'jeg gider ikke have vrøvl med eleverne eller skolen' over 'en lav karakter risikerer en elev mindre til næste år og så kan det betyde klassesammenlægninger og en fyret kollega' til en ' "outguessing" af de kommende eksamenskarakterer', så karaktererne i højere grad afspejler lærerens forventninger til hvordan eleven vil klare sig ved den fremtidige eksamen frem for hvilken præstation, der er ydet i det forgangne år.

Anvendelse af sådanne for karaktergivningen irrelevante vurderingskriterier vil pr. automatik blive afvist som eksisterende af alle involverede parter, og i almindelighed vil det sandsynligvis også være korrekt. Men 'i almindelighed' er ikke tilstrækkeligt. Kun fuldstændig frigørelse

fra disse og lignende irrelevante forhold for karaktergivningen er tilstrækkeligt. Blot en begrænset tillem্পning af en 2 - 3 karakterer efter sådanne kriterier vil være tilstrækkeligt til at sætte hele karaktersystemet under pres og dermed invalidere karakterernes pålidelighed og – hvilket antageligt er meget værre – deres værdi, som beslutningsgrundlag for omgivelserne hvad enten det drejer sig om de videregående uddannelsessteders tildeling af studiepladser eller en kommende arbejdsgivers overvejelser omkring ansættelser.

En eventuel situationstilpasset karaktergivning vil af naturlige årsager blive benægtet af såvel lærere som af skolerne. Naturlige årsager er for det første synonym med, at det er en ren 'tabersag' for alle de primært involverede – i hvert tilfælde så længe alle vender det blinde øje til problemstillingen - idet alle så at sige har affundet sig med at karakterernes 'sandhedsværdi' er ret så begrænset og derfor vil enhver indsats på området være ulønsom. Men en sådan holdning er blot et skridt på vejen mod total værdiløshed og dermed er karaktergivningen en aldeles overflødig og følgelig en frustrerende anstrengelse. For det andet ligger der i 'naturlige årsager', at det dels er moralsk forkasteligt at give karakterer på denne måde og dels en implicit forudsætning om, at det skal være en bevidst (eller stærkere: intentionel) handling. Men det behøver slet ikke at være tilfældet. Selv karakterer afgivet efter lærerens allerbedste overbevisning og helt uden skelen til eksterne forhold kan meget nemt indeholde systematiske afvigelser eksempelvis på grund af en helt ubevidst og efterhånden opstået varemæssig rutineseret karaktergivning – skolerne er også lærende organisationer - men med fuldstændig samme effekt som den bevidste situationstilpassede karaktergivning.

Der kan således peges på mindst to mulige årsager til begrænsninger i karakterernes troværdighed

- manglende viden om systematiske bias i karaktergivningen og
- bevidst såvel som ubevidst tilpasning til 'taktisk karaktergivning'.

Empirisk analyse.

Med henblik på at afdække eventuelle mønstre og strukturer i karakterdannelsen er karaktererne for 18 hhx-klasser fra 5 forskellige skoler analyseret. Analysen omfatter terminskaraktererne første år, årskarakterer første år, terminskarakterer andet år samt årskarakterer andet år for alle basisfag samt valgfagene matematik B og tysk B afgivet efter gældende læseplan – dvs. erhvervsøkonomikaraktererne omfatter både mundtlig og skriftlige karakterer. Med henblik på en vis tværfaglig sammenlignelighed er taget gennemsnittet af skriftlig og mundtlig karakterer for de fag hvori de forekommer. Terminskaraktererne for første år udviste en betydelig spredning og kun begrænset systematik set i forhold til de efterfølgende tre karakterer og er derfor udeladt i nærværende sammenhæng.

Analysemetode

Sigtet med analysen er en afdækning af karaktergivningsmiljøet, som det kommer til udtryk i ovenstående data. Undersøgelsen er i såvel form som indhold eksplorativ og sigter alene mod at afdække objektive målbare forhold som kan antages at influere på karaktergivningen og er altså ikke en vurdering af om karaktererne er korrekte. Med dette som mål beregnedes gennemsnitskaraktererne og karakterspredningen for hvert fag, for samtlige klasser, og samtlige karakterterminer. Herved reduceredes de ca. 10.000 enkeltkarakterer til mere operationelt ca. 400 observationer.

Identificeringen af eventuelle faktors påvirkning af karakterdannelsen er her gjort ved at måle en række forskellige, men plausible faktors samvariation med gennemsnitskarakte-

terne. Der ligger ikke heri noget bevis for kausalitet, men dog en mulig årsag-virkning sammenhæng samt hvilken styrke og retning, den må skønnes at have. Sidstnævnte belyses yderligere ved en opdeling af materialet efter fag og en om end begrænset til tre observationer lang tidsserieanalyse. Herved er det også muligt at identificere eksistensen af eventuelle mønstre og deres faglige og tidsmæssige omfang.

Graden af samvariation er her angivet med korrelationskoefficienten. Den kan antage værdier i intervallet fra -1 til 1 , hvor -1 angiver perfekt negativ samvariation (dvs. hver gang faktor 1 forøges med x enheder falder faktor 2 hver gang med nøjagtig y enheder) medens en værdi omkring 0 viser en manglende samvariation og en værdi på 1 angiver en perfekt positiv samvariation (når faktor 1 stiger med x enheder stiger faktor 2 med y enheder). Værdier herimellem angiver at graden af samvariation er mindre end perfekt – dvs. at når faktor 1 ændres med x enheder kan den tilsvarende ændring i faktor 2 være større eller mindre end y enheder, men korrelationskoefficienten angiver også her den gennemsnitlige styrke og tendens i samvariationen. Bemærk korrelationskoefficienten angiver alene styrken og retningen i samvariationen og ikke effekten af samvariationen – dvs. størrelsen y i de forklarende eksempler¹.

Statistisk set er der 2 interessante – og almindeligt gældende – grænseværdier, nemlig når samvariationen er så udpræget, at den er 95 % eller 99 % sikker – dvs. at der i mindre end 5 % hhv. 1 % af tilfældene ikke er en sammenhæng. De to grænseværdier, der sædvanligvis benævnes signifikant på 5% - hhs. 1%-niveauet, er afhængig af stikprøvens størrelse. Ved en stikprøve på 18 observationer, der er den hyppigst forekommende i nærværende undersøgelse, vil korrelationskoefficienter større end $\pm 0,468$ være signifikante på 5%-niveauet - mao. den fundne samvariation vil forekomme i 95 ud af 100 tilfælde - medens en sikkerhed på 99 % eller derover opnås når korrelations-koefficienten er større end $\pm 0,590$. For en stikprøve på 13 – matematik B og tysk B - er de tilsvarende værdier $\pm 0,553$ og $\pm 0,684$.

Analyse af gennemsnitskarakteren

Gennemsnitskarakteren er et centralt mål for hele klassens præstationsniveau i det pågældende fag og ved at sammenholde udsvingene i de 18 / 13 gennemsnitskarakterer pr. fag med værdierne for en række karakteristiske kendetegn, kan graden af samvariation bestemmes og dermed sandsynliggøre en sammenhæng. Resultatet af beregningerne, der er vist i bilag A, skal kort kommenteres. Af EDB-praktiske årsager benævnes karaktererne 2., 3. og 4. termin for hhs årskarakterer første år, terminskarakterer 2. år og årskarakterer 2. år

Kønsfordeling. En nærliggende antagelse er en samvariation mellem gennemsnitskarakteren og klassens kønsfordeling – her målt som drengenes procentandel. I 70 % af tilfældene er der tale om en positiv sammenhæng, men i intet tilfælde er der tale om en statistisk signifikant samvariation. I Info og Afsætning er der dog en konsistent og relativ høj korrelationskoefficient - mao. klasser med relativt mange drenge har i disse fag en tendens til at have et højere gennemsnit end klasser med relativt mange piger. I de øvrige fag er der en skiftende eller helt ubetydelige samvariation. Om end der er tegn på at drengeskønnet kan slå positivt igennem på gennemsnitskarakteren i visse fag, er den overordnede konklusion, at fordelingen på piger og drenge i en klasse ikke signifikant afspejles i de afgivne karakterer.

Afgangsklassen fra folkeskolen er også en potentiel indflydelsesfaktor og den er undersøgt ved at måle samvariationen mellem gennemsnitskarakteren og andelen af elever udgået fra 10. klasse. I Engelsk kan der konstateres en relativt pæn, men ikke signifikant positiv sammenhæng, medens det er lige omvendt i Tysk. I de øvrige fag er tendensen skiftende over

¹ Hver gang temperaturen stiger 1° (= x enheder) på en celsiuskala er effekten en stigning på $1,8^{\circ}$ (= y enheder) på en fahrenheitskala, men korrelationskoefficienten er 1 , der angiver, at der er tale om en perfekt sammenhæng og at den er positiv.

tiden eller samvariationen er helt betydningsløs hvorfor konklusionen må være at andelen af elever med 10. klasse hhs. andelen med 9. klasse ikke slår igennem til klassens gennemsnitskarakter.

Supplerende uddannelse. Af samtlige elever kom godt 20 % ikke direkte fra folkeskolen, men havde prøvet noget andet før – fra gymnasiet, nogle med HG/EUD-baggrund, nogle HF'ere, højskole, midlertidige jobs mv. Disse elever betegnes 'andel med supplerende uddannelser' og denne andels samvariation med gennemsnitskarakteren er også undersøgt. I Erhvervsøkonomi og Matematik er der en signifikant negativ samvariation mellem klassens gennemsnitskarakter og andelen med supplerende uddannelser i 4. termin understøttet af relativt store koefficienter i 2. og 3. termin. Tendensen er den samme i Info, Historie og Tysk om end i intet tilfælde statistisk signifikant. Af de i alt 24 målinger er viser de 20 en negativ samvariation. Den umiddelbare konklusion vil derfor være, at en supplerende uddannelse eller anden efter-folkeskole-aktivitet udøver en negativ indflydelse på gennemsnitskarakteren og i de fleste tilfælde er den negative samvariation mest udpræget i 4. termin.

Rent logisk forekommer det, at en supplerende / yderligere uddannelse udover folkeskolen skulle påvirke gennemsnitskarakteren i nedadgående retning, at være i modstrid med den almindelige forventning om uddannelsernes effekt. Her er der dog formodentlig tale om en kortslutning i konklusionen, idet faktoren Supplerende uddannelse i sig selv muligvis ikke målbart slår igennem på gennemsnitskarakteren, men at faktorens gennemgående negative samvariation med gennemsnitskarakteren skal forklares med en helt anden mekanisme og via en anden faktor – nemlig fraværet.

Gruppen af elever med supplerende uddannelser har den mest signifikante positive samvariation med gennemsnitlige fravær og som det fremgår af et efterfølgende afsnit, er det en faktor der rykker. Supplerende uddannelses negative samvariation med gennemsnitskarakteren er formodentlig ikke udtryk for at yderligere uddannelse er dårligt eller karaktermæssigt straffes, men da elever med supplerende uddannelser har det markant højeste fravær, vil fraværets negative virkning totalt dominere en eventuel positiv effekt af en supplerende uddannelse.

En anden tolkningsmulighed er, at denne gruppe elever rent faktisk fagligt ligger under gennemsnittet², men at deres supplerende uddannelse overvinder eller skjuler dette forhold de første terminer, og at det så manifesterer sig senere i studiet, jfr. den gennemgående stigende styrke i korrelationen.

Egnethed Den afgivende skoles bedømmelse af eleven er også undersøgt. Her er målt samvariationen mellem gennemsnitskarakteren og andelen af elever med prædikatet Eget. I flere end 85 % af tilfældene er der tale om en positiv samvariation og i 4 tilfælde en statistisk signifikant sammenhæng – mest udpræget i Matematik 4. termin hvor sammenhængen er mere end 99 % sikker, men også i Engelsk og Erhvervsøkonomi er der en klar positiv samvariation. I de øvrige fag er der en tilsvarende tendens om end styrken ikke er så udpræget.

Det synes derfor rimeligt, at konkludere at den afgivende skoles karakteristik Eget slår positivt igennem på klassens gennemsnitskarakter i stort set alle fag. Også her er der dog formodentlig tale om en vis krydseffekt over til fraværsfaktoren, idet gruppen eget har det laveste gennemsnitlige fravær. Det er ikke undersøgt hvor stor samvariationen er ved fastholdt fraværprocent – dvs. Egnethed uafhængig af fraværsfaktoren. En sådan beregning kunne måske ellers vise om Egnethed går på elevens evner eller på elevens ansvarlighed for egen uddannelse eller begge dele.

² Korrigeres der statistisk for fraværet kan der på det individuelle niveau spores en sådan sammenhæng uden den dog er statistisk signifikant – i 4. termin er signifikansniveauet ca. 0,1

Antal elever i klassen. Der er en uomtvistelig positiv samvariation mellem antal elever i klassen og klassens gennemsnitskarakter. I fem tilfælde er korrelationen statistisk signifikant – mest udpræget i Erhvervsøkonomi - og i alt er 20 ud af 24 observationer positive. Denne udbredte positive samvariation må betegnes som ganske uventet, idet den tydeligvis strider mod den almindelige antagelse om forringet undervisningseffektivitet ved øgning af klassekvotienten med deraf følgende forventede lavere gennemsnit i større klasser. Der synes ikke umiddelbart at være mekanismer, som bevirker at klasser med mange elever *skal* have højere gennemsnitskarakterer end små klasser – snarere tværtimod – men dansklærere, engelsklærere, erhvervsøkonomilærere, infolærere, matematiklærere og historielærere værdsætter karaktermæssigt klart større klasser frem for mindre klasser.

Her er der dog formodentlig tale et samspil af flere forskellige faktorer, som forstærker samvariationen over tiden jfr. tendensen med en stigende korrelation over de tre observationer pr. fag. Antal elever i klassen er positivt korreleret med andel Egnede elever, hvilket medfører at store klasser har relativt mange Egnede med opadgående pres på gennemsnitskarakteren. Klassestørrelse er negativt korreleret med fraværet og store klasser vil derfor gennemgående have mindre fravær, hvilket også er positivt for gennemsnitskarakteren.

Der er i andre sammenhænge påvist at store klasser normalt fungerer bedre socialt pga. af relativt færre marginaliserede elever og dette – sammen med en relativ overvægt af Egnede med lavt fravær - er antagelig en medvirkende årsag til at fraværsprocenten i gennemsnit er lavere i store klasser. Endelig kan det ikke afvises, at kausaliteten også går den anden vej – nemlig at høje karakterer og dermed høj gennemsnitskarakter fastholder eleverne i klassen – dvs. formindsker frafaldet.

Fra en anden del af analysen vides, at der kan ses statistisk signifikante forskelle i faglærernes karakterniveau og denne effekt vil utvivlsomt også spille ind her – relativt højt karakterniveau giver højt gennemsnit, der medfører lavere frafaldsprocenter og fordi karakterniveauet er anlagt højt, har store klasser et højt gennemsnit. Konklusionen må derfor nok være, at det ikke er det store antal elever i sig selv, der karaktermæssigt bliver belønnet, men alle de egenskaber og forhold, som medfører, at en klasse har et stort antal elever. Elevantallet må således antages at være en 'proxy'-variabel for andre egenskaber.

Fravær. Som nævnt flere gange er klassens gennemsnitlige fravær klart negativt korreleret med klassens gennemsnitskarakter. I Erhvervsøkonomi er det indiskutabelt den til alle tider mest afgørende karakterpåvirkende faktor af de undersøgte. Dette markante udslag skal sandsynligvis ses på baggrund af, at Erhvervsøkonomi er det eneste helt nye og ukendte fag eleverne møder i skolen med krav om en helt ny og stringent begrebsdannelse og derfor opleves som vanskeligt tilgængeligt. Et fravær vil derfor give nærmest uerstattelige huller i elevernes økonomividen fordi eleverne ikke selv formår at oplæse det tabte under fraværet.

Medens der på dette aggregerede plan ikke er nogen tvivl om fraværets samvariation med klassens gennemsnitskarakter, er denne samvariation slet ikke tilsvarende synlig i den individuelle karaktergivning, hvilket dog ikke fremgår af tallene, men af en anden del af analysen. Forklaringen herpå kan være at sammenhængen er vanskeligere at måle på individniveau, men det kan også tolkes derhen, at klassens gennemsnitlige fravær primært determinerer karakterniveauet – klasseaktiviteten? - og ikke de enkelte karakterer.

Alle fraværsobservationer viser en negativ samvariation med gennemsnitskarakteren bortset fra tysk 4. termin. Tysklærerne udviser i øvrigt en noget egenartet vurdering, der godt nok ikke er signifikant, men dog konsistent. Karaktergivningen i 2. termin svarer til mønstret i de øvrige fag – man værdsætter store klasser og straffer fravær. I 4. termin er det lige modsat – store klasser straffes med et lavere gennemsnit og stort fravær belønnes med højere karakterer! Man foretrækker åbenbart klasser med få elever - også på trods af at det skyldes et lavt fremmøde. Det skal dog bemærkes at tallenes størrelse kun svagt indikerer en sådan konklusion.

De tre sidste faktorer, der undersøges har en noget anden karakter. De er ikke karakterpåvirkende, men belyser nogle strukturelle forhold i karaktergivningen, og som sådan er de i højere grad udtryk for den enkelte lærers tilgang til karaktergivningen. Her undersøges hvordan faglærernes gennemsnitskarakterer samvarierer med 1) spredningen på de af faglæreren afgivne karakterer, 2) klassens totale gennemsnit beregnet som gennemsnittet af elevernes gennemsnitlige karakterer, og 3) den gennemsnitlige afvigelse beregnet som forskellen mellem de af læreren givne karakterer og de forventede karakterer beregnet med udgangspunkt i faglærernes gennemsnitlige vurdering.

Karakterspredningen er udtryk for i hvor høj grad hele karakterskalaen anvendes – dvs. hvor nøje skelnes der mellem de enkelte elevers præstationer. En stor karakterspredning er udtryk for betydelig diskrimination med hensyn til målopfyldelsen medens en lille udtrykker beskeden diskrimination. Ved sammenligning af de enkelte fag bør det her erindres, at i fag med skriftlige karakterer vil der sædvanligvis være et bedre grundlag for at udøve en større diskrimination end i fag med kun mundtlige karakterer.

Bortset fra matematik er samvariationen mellem gennemsnitskarakteren og spredningen på de afgivne karakterer negativ og i rigtig mange tilfælde signifikant negativ, men over tiden en tendens mod svagere samvariation. Dette forekommer aldeles overraskende, idet det implicerer, at jo højere gennemsnitskarakteren er, desto mindre er spredningen på karaktererne. At en sådan sammenhæng kan observeres i enkeltstående tilfælde kan ikke afvises, men den massive negative samvariation viser, at det er regelen snarere end undtagelsen.

Sammenhængen indikerer nemlig at det højere gennemsnit skyldes få eller ingen elever med lave karakterer (hvilket er den spredningsreducerende effekt), men hvis en klasse er et repræsentativt udsnit af befolkningen, vil denne sammenhæng ikke være den generelle – her vil fordelingen være ganske jævn eller eventuelt faldende, dvs. færre elever med høje karakterer – og i så fald vil en højere gennemsnitskarakter være forårsaget af at flere elever ligger højt rent karaktermæssigt, og det medfører øget karakterspredning, dvs. en positiv sammenhæng.

Som nævnt er det den gældende sammenhæng i alle tre terminer i matematik og det skyldes antageligt, at der i Matematik er langt bedre forudsætninger for en meget mere præcis og absolut vurdering af målopfyldelsen og derfor *vil den relative elevandel med lave karakterer være uafhængig af gennemsnitskarakterens højde*, der i stedet påvirkes af andelen med topkarakterer og derfor øges karakterspredningen. Den konstaterede negative sammenhæng kan være forårsaget af mange forhold – eksempelvis en uklar og upræcis målbeskrivelse, indbyrdes uenighed i lærergruppen om hvordan de enkelte delmål skal vægtes eller uenighed om hvad der er essentielt i faget, manglende midler til måling af målopfyldelsen m.fl.

Konklusionen må derfor være, at der er en betydelig relativ karaktergivning i især Dansk, Engelsk og Afsætning, hvilket 13-skalaen bl.a. skulle forsøge at modvirke, idet karakterskalaen forudsættes at være absolut. Den relative karaktergivning medfører nemlig, at en svag elev i en klasse med et højt gennemsnit får en højere karakter end eleven havde fået hvis vedkommende gik i en klasse med et lavt gennemsnit – dvs. elevens karakter er i en vis udstrækning klasseafhængig. Det samme gælder for en stærk elev i en klasse med et svagt gennemsnit – denne elevs karakter bliver mindre her end hvis vedkommende gik i en klasse med højt gennemsnit. Sagt på en anden måde er der altså en tilbøjelighed til at den enkelte lærer fastholder (normal)fordelingens spredning på sine afgivne karakterer, men parallelforskyder kurven efter gennemsnittet. Denne sammenhæng støtter i øvrigt den tidligere fremførte påstand om, at der er statistiske signifikante forskelle mellem de enkelte faglæreres karakterniveau.

Klassens totale gennemsnit. Samvariationen mellem faglærerens gennemsnit og klassens samlede gennemsnit er bortset fra et par enkelte observationer statistisk signifikant. Klassens samlede gennemsnit er alle læreres vurdering af alle klassens elever og er som sådan

nok det mest pålidelige centrale mål for klassens præstationer og evner. I Dansk og Engelsk er samvariationen dog mindre udpræget end i de øvrige fag – mao. har dansk/engelsklærere som gruppe betragtet en vis afvigende opfattelse af klassens præstationer og evner set i forhold til den samlede lærerteams opfattelse. Afvigelsen er ikke statistisk signifikant, men dog synlig.

Gennemsnitlig beregnet afvigelse. Den stærke samvariation mellem disse to størrelser viser blot, at der er en signifikant afhængighed mellem faglærerens karaktergennemsnit og beregnet afvigelse, og at der er tale om en konsistent afvigelse – en 'overnormal' karaktergivning trækker gennemsnittet op og en 'undernormal' trækker snittet ned – helt i overensstemmelse med den positive samvariation mellem gennemsnittet og spredningen.

Analyse af karakterspredningen

Gennemsnittet angiver den centrale tendens, men beskriver kun én dimension af et datamateriale. Den anden dimension er spredningen og den er mindst lige så væsentlig i karaktergivningen. Spørgsmålet er her om der er forhold, der forårsager eller er sammenfaldende med en stor – eller lille - spredning på de afgivne karakterer. På samme måde som i foregående afsnit skal her belyses hvordan spredningen af en classes karakterer udvikler sig i forhold til en række karakteristiske kendetegn

Kønsfordeling. I Historie og Tysk er målt en konsistent og statistisk signifikant positiv samvariation mellem drengandelen i klassen og spredningen på karaktererne – dvs. en sammenhæng, der siger jo større drengandel desto større karakterspredning. Af tallene her kan dog ikke ses om den øgede spredning skyldes at drengene gennemgående har højere eller lavere karakterer end pigerne³ – begge dele vil forøge spredningen. I de øvrige fag er den målte sammenhæng væsentlig mindre om end positiv i flertallet af tilfældene eller med skiftende fortegn over tiden. Konklusionen må derfor være, at medens der ikke kunne måles nogen statistisk sikker effekt af kønsfordelingen på gennemsnitskarakteren, så kan drengandelen i en klasse tilsyneladende have virkning på karakterernes spredning i visse fag.

Afgangsklasse. Bortset fra Info 3. termin er der ingen statistisk signifikante sammenhænge mellem andelen af elever udgået fra 10. klasse og karakterspredningen. Sammenhængen her er negativ og viser at jo større andelen er, desto mindre er karakterspredningen. I flertallet af de øvrige observationer er der også tale om en negativ sammenhæng – i Erhvervsøkonomi konsistent – men styrkemæssigt beskeden samvariation. Konklusionen må altså være, at andelen af 10. klasseselever i en klasse kan have en næsten usynlig dæmpende virkning på karakterspredningen.

Supplerende uddannelse. I matematik og historiefagene er der en konsistent, men ikke signifikant negativ sammenhæng mellem karakterspredningen og 'Supplerende uddannelser'. I de to fag er vurderingen altså den, at mange elever med 'Supplerende uddannelser' i en klasse har en reducerende virkning på karakterspredningen, jfr. her note 2. I de øvrige fag er der tale om en stærk alternerende udvikling og det må derfor konkluderes at andelen af elever med 'supplerende uddannelse' ikke øver nogen betydende systematisk indflydelse på karakterernes spredning.

Egnethed. Andelen af elever med prædikatet 'Egnet' har i flere tilfælde en statistisk signifikant negativ samvariation med karakterspredningen og i de fleste af de øvrige tilfælde er der tale om en betydelig negativ samvariation, dog med en tydelig tendens til aftagende virkning over tiden. Klasser med mange 'Egnede' vil altså have mindre spredning i karaktererne medens få 'Egnede' følges med stor karakterspredning. I Matematik er der en klar positiv samvariation, men her skal der henvises til diskussionen ovenfor vedrørende samvariationen mel-

³ I foregående afsnit påvistes dog at der i gennemsnit er en svag positiv, men ikke statistisk signifikant sammenhæng mellem drengandelen og gennemsnitskarakteren.

lem karaktergennemsnit og karakterspredningen, hvor det konstateredes, at matematik var eneste faggruppe med positiv sammenhæng – mao. mange 'Egnet' udelukker ikke tilstedeværelsen af mindre 'Egnet' i en klasse, men bevirker flere karakterer i den høje ende og derfor større spredning på en classes matematikkarakterer. Korrelationen mellem spredningen og 'Egnet' er altså konsistent med disse observationer. Den overordnede konklusion må altså være at 'Egnet' har en tydelig dæmpende virkning på spredningen af karaktererne.

Fravær. Det gennemsnitlige fravær angiver i et tal den centrale tendens for en hel classes fravær, men når interessen vedrører spredningen af karaktererne – dvs. individualiseringen – er det mere plausibelt, at spredningen på fraværet vil være et bedre mål og derfor er samvariationen for begge medtaget i analysen.

Generelt er der en positiv samvariation mellem karakterspredningen og det gennemsnitlige fravær, men i Info 2. termin er der en meget stærk positiv sammenhæng – dvs. jo større fravær, jo større er spredningen af infolærernes karakterer. I og med at Info første år i vid udstrækning er et færdighedsfag og at tilegnelse af denne færdighed opnås gennem øvelser, der kræver tilstedeværelse, er det rationelt, at der vil være en stor positiv samvariation. I Matematik er der en tydelig negativ, men ikke signifikant sammenhæng, som understøtter den tidligere skitserede forklaring på karakterdannelsen i dette fag: stigning i fraværet reducerer karaktererne og de lavere karakterer resulterer i mindre spredning af karaktererne.

Spredningen i elevernes fravær udviser i store træk det samme mønster som det gennemsnitlige fravær med et par undtagelser i Matematik og Historie med signifikant negativ samvariation. Det er i øvrigt bemærkelsesværdigt at spredningen på fraværet ikke slår hårdere igennem på karakterspredningen. Isoleret set kunne det forventes, at en stor spredning i elevernes fravær ville give anledning til stor spredning i karaktererne, men det synes ikke at være tilfældet. En mulig forklaring herpå kan være at nogle lærere giver udpræget standpunktskarakterer – hvilken målopfyldelse *kan* eleven præstere her og nu - medens andre giver præstationskarakterer – hvilken indsats og hvilken målopfyldelse *er opnået* i forgangne periode⁴. En anden mulig forklaring kan være, at den virkelige sammenhæng skjules af modgående og krydsende tendenser i andre faktorer.

Klassens samlede karaktergennemsnit. Som nævnt under fravær er spredningen på det samlede gennemsnit antagelig mere interessant og derfor indgår de begge i analysen. I almindelighed er klassens samlede gennemsnit negativt korreleret med spredningen på fagkaraktererne således at et højt gennemsnit findes sammen med en lav spredning. Denne sammenhæng er afledt af den tidligere diskuterede negative sammenhæng mellem fagkarakterernes gennemsnit og deres spredning. Derimod er der en entydig positiv samvariation mellem spredningen på fagkaraktererne og spredningen af elevernes gennemsnitlige karakterer, hvilket viser, at der på tværs af lærergrupperne er en betydelig enighed omkring den relative individualisering af karaktererne.

Gennemsnitlige afvigelse. Den beregnede samvariation er udpræget negativ – dvs. stor spredning i de afgivne karakterer gennemgående er sammenfaldende med lille afvigelse fra de forventede karakterer medens en lille karakterspredning typisk ledsages af større afvigelser fra det gennemsnitlige forventede. Der er således tale om en konsistensmåling, der bekræfter niveauforskelle i lærernes karakterafgivning.

Sammenfattende bemærkninger

Sigtet med undersøgelsen har været en belysning af nogle strukturelle forhold i og omkring karakterdannelsen. Der er påvist en række signifikante sammenhænge og mønstre, hvoraf de fleste må betegnes som værende i overensstemmelse med det forventede, men også fle-

⁴ Denne konklusion understøttes af det forhold, at der på det individuelle niveau i mange tilfælde ikke kan spores nogen effekt af elevernes fravær.

re som både er uventede og til dels direkte irrationelle. Blandt de uventede var 'supplerende uddannelses' negative og antal elever pr. classes positive samvariation med karaktergennemsnittet, men baggrunden herfor skal antagelig søges i, at de to variable dækker over – er proxy for - helt andre egenskaber end den pålydende værdi. Fraværets negative samvariation med gennemsnitskarakteren var forventelig, men mere uventet var den forholdsvis begrænsede samvariation på det individuelle plan. Den udprægede negative sammenhæng mellem gennemsnitskarakteren og spredningen på karaktererne synes derimod ikke at kunne begrundes rationelt.

Set i relation til den grundlæggende antagelse ved karaktergivningen – at den er lærer- og situationsuafhængig samt absolut – kan der således stilles et begrundet stort spørgsmål ved denne forudsætnings holdbarhed. En stor klasse med mange egnede og få med anden efter folkeskoleaktivitet og gerne med en overvægt af drenge synes at være en ret sikker vej til en høj gennemsnitskarakter for alle.

Faggruppemæssigt har erhvervsøkonomilærergruppen den mest konsistente karakterdannelse - og det gælder både over tiden og over de forskellige variabler - medens dansklærergruppen har den mest ujævne karakterdannelse blandt fagene med skriftlige karakterer. Baggrunden herfor er ikke søgt belyst. Der er dog ingen tvivl om at lærerforhold som f.eks. uddannelse, alder, erfaring også vil være potentielle faktorer til forklaring af variationer i karaktergivningen.

Det forhold at en samvariation med stor sandsynlighed eksisterer, er dog ikke i sig selv ensbetydende med at de pågældende variabler - uanset hvor stærk samvariationen måtte være - har en stor og afgørende indflydelse på karakterdannelsen. Dels kan effekten være marginal⁵, dels implicerer den ikke i sig selv nogen kausalitet. Som minimum for kausalitet må der også være en rationel begrundelse for eller mekanisme som betinger sammenhængen, og i den udstrækning en sådan forekommer plausibel, er den anført. På det operationelle plan er det selvfølgelig ikke nødvendigt med en sådan begrundelse – her synes en kraftig indsats mod fravær, en maximal spredning udover klasserne af elever med supplerende uddannelse og bestræbelser mod en mere homogen karakterdannelse i alle tilfælde at være en god idé.

⁵ I en regressionsmodels termer svarer det til at β -koefficienten er relativ beskedent

Bilag A: Observerbare faktors samvariation med gennemsnitskarakter

Faktorer	Dnsk	Dnsk	Dnsk	Eng	Eng	Eng	Eøk	Eøk	Eøk	Info	Info	Info
Dreng%	-,070	,061	,102	,064	-,056	-,087	,089	,050	,071	,309	,207	,465
10klasse%	,113	,102	-,018	,402	,146	,325	-,054	-,076	-,056	,284	,224	-,179
Suppl.udd%	-,146	,035	-,058	-,195	,048	,017	-,331	-,386	-,511	-,168	-,254	-,365
Egnet%	,174	,022	,041	,518	,069	,211	,470	,374	,474	,298	,297	,218
Antal elever	,185	,342	,364	,440	,220	,444	,130	,614	,621	,306	,235	,477
Gns. fravær	-,348	-,052	-,206	-,311	-,179	-,267	-,601	-,637	-,668	-,400	-,384	-,483
Sprd(fag)	-,512	-,356	-,476	-,442	-,541	-,512	-,208	-,183	-,445	-,367	-,355	-,115
Klassegns	,314	,469	,525	,556	,575	,540	,799	,816	,893	,583	,462	,653
Gns. afv.	,803	,799	,809	,776	,829	,889	,692	,713	,848	,704	,843	,819
Antal obs.	18	18	18	18	18	18	18	18	18	18	18	18

													Resumé gns		
	Mat	Mat	Mat	Tysk	Tysk	Tysk	S&S	Afs	Afs	Samf	Hist	Hist	Gns	Neg	Neg%
Dreng%	-,464	,015	-,052	,018	-,106	,109	,162	,054	,420	,151	-,233	,129	0,073	7	0,29
10klasse%	-,005	,015	,184	-,098	-,517	-,294	-,019	-,220	,031	,044	-,085	,024	0,023	12	0,50
Suppl.udd%	-,395	-,194	-,582	-,134	-,117	-,088	-,130	,052	-,277	-,073	-,292	-,406	-0,202	20	0,83
Egnet%	,441	,327	,790	,229	-,195	-,002	,248	-,163	,220	,276	,230	,250	0,231	3	0,13
Antal elever	,132	,503	,654	,225	-,045	-,219	-,050	,195	,141	-,066	,519	,519	0,290	4	0,17
Gns. fravær	-,515	-,193	-,812	-,571	-,154	,158	-,352	-,061	-,179	-,443	-,184	-,423	-0,346	23	0,96
Sprd(fag)	,278	,248	,157	-,705	-,168	-,572	-,782	-,518	-,211	-,556	-,392	-,290	-0,349	21	0,88
Klassegns	,887	,804	,786	,822	,734	,681	,652	,784	,695	,603	,792	,858	0,669	0	0,00
Gns. afv.	,675	,571	-,027	,818	,706	,636	,780	,852	,665	,699	,853	,360	0,723	1	0,04
Antal obs.	13	13	13	13	13	13	18	18	18	18*)	13	18			

*) En skole kunne ikke finde terminskarakterer i historie

Årskarakter 1. år
 Terminskarakter 2. år
 Årskarakter 2. år

Signifikans:

	18 obs	13 obs
5%	± ,468	± ,553
1%	± ,590	± ,684

Bilag B: Observerbare faktors samvariation med karakterspredningen

	Dnsk	Dnsk	Dnsk	Eng	Eng	Eng	Eøk	Eøk	Eøk	Info	Info	Info
Dreng%	-,018	,026	,137	,136	,155	-,010	-,143	,266	,046	,066	-,189	-,009
10klasse%	-,204	,042	-,207	,037	-,011	-,144	-,037	-,062	-,018	,012	-,534	-,138
Suppl.udd%	,034	-,182	-,142	,238	,088	-,117	,300	-,350	,018	,354	-,200	-,065
Egnet%	-,249	-,312	-,219	-,404	-,446	-,401	-,498	-,054	-,245	-,568	-,038	-,193
Gns. fravær	,237	,045	,005	,413	,413	,385	,466	,019	,220	,707	,049	,310
Sprd(frav)	,466	,261	,134	,386	,311	,284	,417	-,101	,025	,744	-,156	,247
Klassegns	-,296	,052	-,005	-,491	,003	-,320	-,031	-,204	-,335	-,424	-,348	-,060
Sprd(klgn)	,158	,184	,610	,596	,305	,460	,290	,800	,755	,450	,565	,038
Gns. afv.	-,328	-,436	-,550	-,175	-,712	-,450	-,309	-,084	-,467	,045	-,212	-,060
Antal obs.	18	18	18	18	18	18	18	18	18	18	18	18

													Resumé sprd.		
	Mat	Mat	Mat	Tysk	Tysk	Tysk	S&S	Afs	Afs	Samf	Hist	Hist	Gns	Neg	Neg%
Dreng%	,238	,263	-,016	,160	,363	,713	-,056	,165	,303	,211	,557	,593	0,151	7	0,29
10klasse%	-,233	-,018	,102	-,228	-,003	,400	-,163	,276	,017	,017	-,154	-,041	-0,057	16	0,67
Suppl.udd%	-,206	-,363	-,232	,165	-,062	-,137	,239	-,077	,234	-,304	-,332	-,336	-0,050	15	0,63
Egnet%	,254	,391	,164	-,400	-,071	,223	-,502	,043	-,377	,056	,357	,147	-0,163	16	0,67
Gns. fravær	-,320	-,346	-,274	,504	-,393	-,294	,522	,084	,466	-,067	-,269	-,064	0,145	8	0,33
Sprd(frav)	-,279	-,554	,130	,431	-,042	-,204	,270	-,175	,380	-,131	-,610	-,317	0,101	10	0,42
Klassegns	,340	-,032	,623	-,603	,154	-,013	-,668	-,340	-,447	-,232	-,419	-,208	-0,196	19	0,79
Sprd(klgn)	,598	,529	,256	,541	,398	,165	,436	,223	,346	,673	,635	,424	0,434	0	0,00
Gns. afv.	,037	,459	-,828	-,549	-,381	-,896	-,490	-,507	,145	-,443	-,202	-,162	-0,313	20	0,83
Antal obs.	13	13	13	13	13	13	18	18	18	18*)	13	18			

*) En skole kunne ikke finde terminskarakterer i historie

Årskarakter 1. år
 Terminskarakter 2. år
 Årskarakter 2. år

Signifikans:

	18 obs	13 obs
5%	± ,468	± ,553
1%	± ,590	± ,684

Ja - hvor tit anvender man egentlig den laveste dumpekarakter?

Med en række nærmest ubefæstede påstande og afsluttende med en retorisk finte annoncerer prorektor for Århus Universitet og formand for Karakterkommissionen Katherine Richardson en ny karakterskala i Jyllands Postens kronik den 12.12. – af kommissionen og undervisningsministeren døbt 12-skalaen.

Udgangspunktet for prorektorens kronik - og Karakterkommissionen - er at 13-skalaen ikke er internationalt kompatibel! – med udråbstegn. Ja det må jo siges at være en særdeles alvorlig mangel ved den karakterskala, der har målt og vejet alle danskere i de seneste 40 år, men hvor mange andre landes karakterskalaer er det nu der er internationalt compatible? Bortset fra Norge, og her med særderles dårlige erfaringer - ingen! At der er tale om en væsentlig egenskab ved en karakterskala forsøges argumenteret hjem med at det er en skam (prorektorens ord), at mange dygtige danske studerende, som et led i deres forberedelse på at træde ind i det globale arbejdsmarked, tager en del af deres uddannelse i udlandet og at disse unge mennesker uddannet i andre systemer ingen muligheder har for at få overført de karakterer, de optjener i udlandet, til et dansk eksamensbevis. Det kan bare ikke lade sig gøre, så længe vi opererer med den nugældende karakterskala. Jamen prorektoren og kommissionen har ret, men hvor mange er det nu det egentlig, der har brug herfor? og løser den nye skala dette problem? Nej vel.

Kommissionen hænger godt nok sin nye karakterskala op på den såkaldte ECTS-skala – en monstrøs uigennemsigtig naturstridig skala, der bedst kan sammenlignes med et gummibånd uden ende og uden midtpunkt – altså et aldeles ubrugeligt målebånd. Den er afhængig af det til enhver tid værende underviste segment og ikke af den opnåede viden og indsigt i faget. Det er kommissionen selvfølgelig helt klar over og forsøger derfor med forskellige akademiske kunstgreb at forbedre den danske udgave bl.a. ved at kræve at skalaen skal være absolut. Hvordan man med nogen sikkerhed kan oversætte fra en fast skala (alle andre end den norske karakterskala) til elastisk skala –ECTS-skalaen - og tilbage fra en elastisk skala til en fast dansk 12-tals skala uden tab af informationer – ja se det mangler kommissionen godt nok at forklare.

Men det er kun den ene side af problemstillingen. Den anden er, at den karakterafgivende institution også anvender en lignende kompatibel skala, der kan konverteres til ECTS-skalaen. Intet tyder for nærværende på at de højere lærestalter i de engelsktalende lande – England, USA, Canada og Australien – generelt eller for den sags skyld specielt vil anvende en sådan skala. Med de stærke traditioner i den franske undervisningssektor in mente og den til tider hidsige karakterdebat i Sverige under udviklingen af den nugældende 4-punkts-skala er der heller ikke nogen begrundet formodning om at disse lande umiddelbart står overfor at indføre en ECTS-kompatibel skala. Med et slag på tasken vil ca. 80 % af de unge danske studerende i udlandet frekventere højere lærestalter i disse lande – og hvad vil de så få ud af at Danmark har en kompatibel skala? Intet – i hvert tilfælde ikke i de næste mange mange år, så karakterimporten vil næppe forøges væsentlig med den nye skala foreløbig. Den opsamlede viden på de udenlandske lærestalter kan naturligvis fint importeres – den er ikke afhængig af en karakterskala!

Ifølge kommissionsformanden er et andet væsentligt argument mod 13-skalaen at det er umuligt at eksportere dygtighed i denne indpakning fordi udlændinge ikke har forståelse for skalaens 13-tal. Argumentet er velkendt – især fra amerikanske lærestalter, men en korrekt beskrivelse af 13-skalaen fjerner som regel hurtigt denne misopfattelse. 13-skalaen kan i modsætning til alle an-

dre karakterskalaer entydigt og i et for stort set alle – i det mindste med en højere uddannelse - forståeligt sprog beskrives kort og fyndigt - med omtrentlige parametre - sådan $N(8, 1,5)$ eller verbalt formuleret en normalfordelt skala med et i runde tal gennemsnit på 8 og en standardafvigelse på 1,5 og det er et sprog især de amerikanske læreanstalter forstår. Så mon ikke det i højere grad er tale om forklaringsproblem end et forståelsesproblem?

Manglende international kompatibilitet var karakterkommissionen væsentligste begrundelse for at anbefale en afskaffelse af 13-skalaen, men ikke den eneste. En anden væsentlig kritik af 13-skalaen går på, at der har udviklet sig forskellige traditioner for dens anvendelse indenfor forskellige fag og forskellige uddannelser. Som dokumentation for den faglige forskellighed i anvendelsen af 13-skalaen vises i betænkningen en grafisk fremstilling af karakterfordelingen i skriftlig dansk og skriftlig matematik i hf for den samme elevgruppe. Grafen viser det velkendte forhold at matematikkaraktererne følger kamelfordelingen med 2 pukler og dansk karaktererne følger dromedarfordelingen med 1 pukkel. Et nærliggende spørgsmål i denne forbindelse ville naturligvis være om der er tale om dårlige måleredskaber eller dårlige målinger eller om der er tale om en Ole Rømer-måling af et – endnu? - ukendt fænomen.

Som dokumentation for forskelligheder uddannelser imellem er gengivet 3 undersøgelser af Inge Henningsen, Københavns Universitet. Umiddelbart er der ikke grund til at pege fingre af de 3 analyser, men man kan med rimelighed stille spørgsmål ved analysernes validitet i relation til Karakterkommissionens opgave. Det gør forskeren da også selv med "..... at det ikke er muligt i modellen at skille brug af karakterskalaen fra andre faktorer.....". Kommissionen accepterer dog uden videre de konstaterede karakterforskelle ved helt at negligere betydningen af *andre faktorer*, men det ville nok have været betydeligt mere givtigt at undersøge hvorfor forskellene manifesterer sig. Det er altså svært at fremkomme med en holdbar løsning uden at kende årsagen til problemet!

Det er næppe 13-skalaen, der er *årsagen* til disse uønskede tilstande fagene og studierne imellem, men at rette smed for bager er jo en århundredgammel tradition i Danmark – ja endda ældre end den berømmelige H.C. Ørsted skala.

Efter karakterkommissionen mening kan disse uheldige egenskaber ved 13-skalaen elimineres med en ny skala baseret på ECTS-skalaen og som i absolutte termer skal måle *graden af målopfyldelse*. Rent parentetisk kunne man så spørge om hvad var det nu var 13-skalaen målte? Den manglende kompatibilitet mellem ECTS-elastikken og en dansk skala med absolutte vurderinger snor kommissionen sig rundt om ved at opstille følgende forudsætninger, nemlig at:

- målene for faget
- prøven
- populationen

skal opfylder følgende to centrale antagelser.

Antagelse 1

Målene for et fag afpasses – i det mindste over tid – til den del af population, der består prøven i pågældende fag.

Antagelse 2

Ved tilrettelæggelsen af en prøve tilstræbes det gennem en afpasning af sværhedsgraden, at resultaterne af de beståede

præstationer vil være nogenlunde ligeligt fordelt inden for det samlede interval af mulige beståede præstationer.

Hvorvidt disse forudsætninger er realistiske eller kan antages at være realistiske diskuteres overhovedet ikke af kommissionen. Såfremt disse forhold ikke er til stede i den virkelige verden – altså ude i undervisnings- og prøvesituationen - må de betegnes som rene akademiske mentalakrobatiske øvelser. Den første antagelse må betegnes som en rigtig devalueringssklausul. Når den danske krone devalueres sker der ikke noget med danske penge – en hundredkroneseddel er stadig en hundredkroneseddel efter devalueringen. Det er først når den skal omveksles til udenlandske varer man opdager effekten. Den samme effekt vil være resultatet af den første antagelse – et 10-tal vil stadig være et 10-tal efter en nedskrivning af målene, men en arbejdsgiver vil hurtigt mærke, at de 2 10-taller ikke har samme værdi. En måltilpasning af et fag for at sikre, at tilstrækkelig mange består eller opnår et 12-tal, har den naturlige virkning at hæve beståelsesprocenten og sænke uddannelsesniveaet - en erfaring mange undervisere kan bevidne med udviklingen gennem de seneste ca. 15 - 20 år.

Især antagelse 2 er der alvorlige og praktiske grunde til at afvise. Med den eksisterende 13-skala er der en tilnærmelsesvis 1:1 sammenhæng mellem populationen og karaktererne, idet begge stort set er normalfordelte, men den sammenhæng vil der ikke være med den foreslåede 12-skala. Konstruktion af sådanne prøver med afpasning af sværhedsgrader, der sikrer en ligelig fordeling indenfor beståelsesrammen af den beståede andel af populationen, vil ligge langt uden for de erfaringer der er til stede i hele den danske undervisningssektor og den kognitive forskning vil næppe heller kunne forbedre den situation.

De anførte forudsætninger er absolut nødvendige, men de er ikke tilstrækkelige. De kan nemlig ikke sikre at et 7-tal er et 7-tal til alle tider og i alle sammenhænge – dvs. at karakterskalaen er absolut - og det var netop kommissionens andet argument for at forkaste 13-skalaen. For at kunne give en karakter, der afspejler graden af målopfyldelse, må der foreligge en – i teorien additiv – målbeskrivelse for *faget*, der sædvanligvis angives i en skala fra 1 til 100. For at kunne anvende ECTS-skalaen skal der også fastlægges den grad af målopfyldelse, der er nødvendig for at bestå – eksempelvis 30 % af fagets målbeskrivelse. ECTS-skalaen beskriver med bogstaverne A til E til kun de 70 % der *består*, og her er kravet at

- de øverste 10 % af de beståede får A,
- de næste 25 % får B,
- efterfølgende 30 % får C,
- de næste 25 % får D og
- de 10 % dårligst placerede af de beståede tildes et E.

Fordelingen af de beståede er pæn og nydelig symmetrisk omkring C, men hvordan er evnerne fordelt hos de unge - elever og studerende - set i forhold hertil? Det er jo trods alt den virkelighed, som skalaen skal måle. Alle studier synes at bekræfte at evnerne kan beskrives med en normalfordeling og den er jo også pæn og nydelig symmetrisk, *men den omfatter også de ikke-beståede efter ECTS-skalaen*. Midtpunktet for fordelingen af *evnerne* ligger altså noget lavere end karakterskalaens midtpunkt.

En simulering af disse betingelser for et fiktivt fag med lineær målbeskrivelse viser meget hurtigt at de af kommissionen opstillede betingelser ikke er tilstrækkelige. For det første - hvis fordelingen af studerende til forskellige studier er forskellig - holder karaktererne ikke. Eksempelvis hvis alle "genierne" søger ind på ingeniørstudiet og alle "tåberne" søger ind på fysikstudiet vil de resultere

rende karakterer for samme fag ikke afspejle de studerendes viden – "tåberne" vil have meget nemmere ved at få 10 end "genierne" – eller sagt med andre ord: "genierne" vil opleve et markant karakterfald og "tåberne" vil opleve en tydelig karakterstigning i forhold til deres udgangspunkt. Her vil kommissionen naturligvis påpege at dens forudsætninger brydes, men det er egentlig ligegyldigt for det er den virkelighed, der eksisterer. Ved mange studier er der adgangskrav, som kun gør det muligt for "genier" at komme ind, medens der ikke er adgangskrav på andre studier og her vil "tåberne" have muligheder.

For det andet vil forskellige beståelsesgrænser også gøre karaktererne invalide – eksempelvis hvis lægestudiet kræver at bestået anatomi svarer til 50% af fagets mål medens tandlægestudiet kun kræver 30% målopfyldelse vil samme karakter heller ikke afspejle de studerendes viden på de 2 studier.

For det tredje skal målbeskrivelsen for de enkelte fag også være identisk – beskrives dansk lineært og matematik eksponentielt vil et 10-tal i dansk afspejle en anden viden end et 10-tal i matematik. Det er muligvis dette forhold, der ligger til grund for den konstaterede forskel mellem engelsk og matematik karaktererne i Inge Henningsen første analyse. Endelig viser simuleringen at jo højere man kører beståelsesgrænsen op desto mere fintmasket – og dermed kompleks – skal prøven være. Det kan derfor fastslås at udover de af kommissionen opstillede forudsætninger må der yderligere – mindst – stilles et krav om ens og konstante beståelseskrav for samtlige fag og alle institutioner og en fuldstændig ensartet målbeskrivelse på tværs af samtlige fag.

Men det tekniske fundamentet for 12-skalaen er ikke det eneste sted man kan sætte en finger på. Sædvanligvis anses en karakterskala for at have 2 væsentlige funktioner nemlig den eksterne – institutionens afrapportering til omgivelserne om hvilken mængde viden den pågældende person har erhvervet – og den interne – styringen af personens bestræbelser under studierne. Den interne funktion – ofte benævnt den formative vurdering - er således primært et anliggende mellem underviseren og den underviste, hvor underviseren bl.a. ved hjælp af karakterskalaen kan vejlede og vurdere den undervistes indsats ud fra en for begge parter kendt og anerkendt målestok. Det siger sig selv, at for en underviser er den formative funktion den absolut væsentligste både i den løbende evaluering og ved afgivelse af års- og terminskarakterer.

Karakterkommissionen behandler den formative funktion meget stedmoderligt, idet den er indskrænket til krav om at en karakterskala skal have flere karaktertrin for at kunne anvendes til dette formål, og det har 12-skalaen. Men skalaens indretning gør den dog nærmest uanvendelig til dette formål, idet den er så grovkornet, at den er intetsigende i relation til de observerbare ændringer i indsatsen. Er beståelseskravet for et givet fag eksempelvis sat ved en målopfyldelsesgrad på 30 vil 12-skalaens karakter på 7 dække intervallet fra en målopfyldelsesgrad på 54,5 til 75,5. En elev der starter med en målopfyldelse på 55 vil således opnå et 7-tal. Med en ihærdig indsats både hjemme og i undervisningssituationen lykkes det at nå op en målopfyldelse på 75 til afsluttende eksamen og eleven får stadigvæk et 7-tal! En indsats der er forbedret med ca. 35 % vil altså ikke have nogen effekt på 12-skalaen. Det må da vist være enhver undervisers værste mareridt. Det hjælper ikke underviseren – som kommissionen gør det - med at henvise til at i den løbende evaluering, kan han fint anvende de mellemliggende tal, for både 6 og 8 kommer ud som et 7-tal på eksamensbeviset.

Det næstvæsentligste argument mod 13-skalaen er ifølge kommissionen, at den afbilder dansk og matematik forskellig – matematik som en kamelfordeling og dansk som en dromedarfordeling. Lø-

ser 12-skalaen dette problem? Endnu har vi ikke erfaringer at bygge på, men alt taler for at den gør det. Men hvordan sluger kommissionen den kamel? – på strudsemaner ved at gøre skalaen så grovkornet at den bliver usynlig! Ikke et eneste argument for at fordelingen er forkert eller fejlagtig i forhold til virkeligheden. Men hvis det nu *er* virkeligheden at evnen til at løse matematiske problemer deler eleverne i 2 grupper så mistes der værdifuld information ved at anvende 12-skalaen. På atletikbanen deler stopuret også løberne op i forskellige grupper, men derfor kan alle løbere sagtens være lige gode eller lige dårlige til svømning.

Også på et tredje område vil 12-skalaen grovkornethed antagelig få betydning. Kommissionen lægger afgørende vægt på at en dansk karakterskala skal gøre det muligt at beregne et gennemsnit, bl.a. fordi der er tradition for at anvende gennemsnittet i forskellige sammenhænge som beståelses- og adgangskrav. Med de store spring i karaktererne vil det med meget stor sandsynlighed – og i fin overensstemmelse med undervisningsministerens ønske - medføre en tydelig større beståelsesprocent sammenlignet med 13-skalaens beståelsesprocent.

Så hvor stort et behov er der egentligt for argumentet med en negativ dumpekarakter for at forkaste 12-skalaen?

Bilag a: Citater fra betænkningen

Formålet med nedsættelsen af kommissionen er at undersøge om den nuværende anvendelse af karaktersystemer giver et retvisende og gennemsigtigt billede af den individuelle bedømmelse af elever og studerende i uddannelsessystemet.

P 6

Det er en forudsætning for kommissionens arbejde, at karakterskalaen skal kunne udtrykke graden af målopfyldelse med hensyn til faglig viden. Dette skal også ses i sammenhæng med den indholdsudvikling, der gennemføres i det danske uddannelsessystem som led i regeringens handlingsplan "Bedre Uddannelser" fra 2002.

P 7

overgangen fra stram indholdsstyring til målstyring, herunder at det enkelte fag ikke længere skal beskrives ud fra en række indholdselementer, men ud fra klare, evaluerbare mål

P 11

Regeringen sætter i denne redegørelse fokus på, at danske uddannelser skal kunne måle sig med de bedste i verden, og at danske forskeruddannelser lever op til den højeste internationale standard. Hjerneflugt kan undgås, ved at danske uddannelser er så tidssvarende og attraktive som muligt. Samtidig skal det sikres, at uddannelserne kvalificerer danske elever, studerende og arbejdstagere til at klare sig i internationale miljøer.

P 12

Udgangspunktet for den danske karaktergivning er den absolutte bedømmelse. Men før den endelige karakter fastlægges, skeles der ved nogle prøveformer til den opnåede karakterfordeling, hvorefter den første foreløbige karaktergivning eventuelt justeres. Det er oftest ved skriftlige prøver med deltagelse af store populationer, at karaktergivningen foretages i disse to trin.

P 14

Denne slitage kan for eksempeludmønte sig i, at kun en begrænset del af skalaen benyttes, at der over tid iagttages en karakterglidning (karakterinflation), eller at der udvikler sig en konsensus fra fag til fag og fra uddannelse til uddannelse om, hvordan karakterfordelingen bør være.

P 15

Samtidigt hermed kan der ske ændringer i de faglige krav, der stilles (videnglidning). Det kan ikke udelukkes, at en lang periode med meget ringe karakterer ved en bestemt prøve i et fag fører til en lempelse af de faglige krav, ligesom en generel højnelse af eksaminandernes viden og kunnen i et fag kan føre til stramninger af de faglige krav.

P 15