

Ministeriet for Videnskab,
Teknologi og Udvikling
Bredgade 43
1260 København K.
Att: Annette Rasmussen
Udvalget vedrørende
Videnskabelig Uredelighed

UBST	Sagsbehandler ERC
Sags nr. 75047	Dok.id. 481302
DATO 20/9.06	Bilag:

2

Adsløvet den 18. september 2006.09.17

Kære Annette Rasmussen.

Det slog mig at udvalgsarbejdet måske kunne lettes gennem fremsendelsen af en overordnet sagsfremstilling - med mine kommentarer til relevante passager i dekanens indberetning til rektor (bilag 3) – en uddybende og måske mere forståelig version af bilag 4.

Jeg skal derfor bede om at vedhæftede brev indgår som bilag 27 til mit brev af 14. ds. til UVVU.

Med venlig hilsen
Helmuth Nyborg



Overordnet sagsfremstilling (delvist på engelsk).

Indledning

I dekan Svend Hyllebergs tjenstlige indberetning af professor Helmuth Nyborg, dateret 7. august 2006 (SAM-2005-602-001) anfører dekanen bl.a., at han vil "... indberette forhold, der efter [hans] vurdering udgør en alvorlig forseelse mod Helmuth Nyborgs pligter som professor ved Aarhus Universitet." Ligeledes anføres, at "I forbindelse med forskningsprojektet "Sex-related differences in general intelligence" har Helmuth Nyborg ikke levet op til fundamentale krav om videnskabelig redelighed og god forskningsetik".

Mere specifikt anføres, at "... han har efter min vurdering udvist en groft uagtsom adfærd, som indebærer en utilbørlig vildledning om egen videnskabelig indsats og resultater. Dette er sket ved at publicere resultater fra en undersøgelse, der er baseret på en helt utilstrækkelig beskrivelse af undersøgelsens design, en meget mangelfuld monitorering af undersøgelsen, en beskrivelse af grundlaget for de offentliggjorte undersøgelser, der ikke blot er mangelfuldt, men også flere steder ukorrekt, samt dokumenteret et helt utilstrækkeligt kendskab til de anvendte statistiske metoders anvendelighed og begrænsning. Helmuth Nyborg har således, som professor ved Aarhus Universitet, publiceret en helt utroværdig og værdiløs undersøgelse, og ved sin forsømmelighed forbrudt sig alvorligt mod kravene til videnskabelig redelighed og god forskningsskik.

Endvidere anfører dekanen, at "Udover dette har Helmuth Nyborg i forbindelse med fakultetets undersøgelser også forbrudt sig mod sin loyalitetsforpligtelse over for Aarhus Universitet ved at rundsende skrivelser, hvor han gennem urigtige oplysninger stiller universitetet i et urimeligt dårligt lys."

I det følgende vil jeg først kort præsentere sagen som jeg ser den (på engelsk og dansk). Dernæst kommenteres de 10 punkter dekanen anfører som dokumentation for min groft forsømmelige adfærd. Endelig kommenteres de påståede brud mod loyalitetsforpligtelsen.

Sagen som jeg ser den

Objektivt ser sagen således ud for mig.

In 2001 and 2002 I discussed various methodological aspects of the observation that sex loads .272 on general intelligence g in adults at the annual meetings of the International Society for Intelligence Research (ISIR2001).

In 2003 I published the result in chapter 10, Sex differences in g , in H. Nyborg (ed.) *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, Oxford: Pergamon, in addition to discussing methodological problems and the results of others.

In 2005 I finally published a slightly modified version of the finding in H. Nyborg, Sex-related differences in general intelligence g , brain size, and social status in the respected peer-reviewed journal *Personality and Individual Differences*, 39, 497-509.

The number of adult subjects with full data went up from 52 to 62 over the period in the ongoing project, and after correction for a previous minor error in calculation, sex loaded in the final paper .274 on g - a hardly noticeable change from the original ISIR2001 finding of .272. The nature and consequences of the error in calculation was spelled out at the public home page of the 2003 book (Elsevier).

The hierarchical factor-analytic method was used throughout the studies and the test for sex differences followed the recommendations by the leading authorities in the field (e.g. Jensen, 1998, p. 538), i.e. to examine whether sex loads significantly on g . The result did not deviate from what other researchers had found, i.e. a 3-5 IQ point sex difference, but it was noted that some researcher do not find a difference. The paper discussed the methodology and the likely reasons for discrepancies.

More generally, over the past 35 years I have worked within a general scientific frame in order to produce books and articles, covering a number of widely differing research areas, in professional journals. Most of the international journals are "high-impact", and my average citation frequency is above 9. Reviews have generally been quite positive. I believe my approach is fully in line with how my international peers proceed in scientific matters. Within the past decennium I have edited two scientific books of 600+ pages, each, published on a prestigious printing house; I have authored or co-authored numerous articles in *Intelligence, Personality and Individual Differences*, and have recently submitted more articles. I have lectured at countless international conferences over the years, generally with favourable feedback.

However, the moment I first reported on a modest sex difference in general intelligence, *g*, I was claimed to demonstrate a marked lack of insight into data and methodology in a field I have been quite active within for years, and dean Svend Hylleberg reported to the rector of the University of Aarhus that I have produced a completely untrustworthy and worthless investigation. I find this treatment unfair for a number of reasons given below.

Kommentarer til de 10 punkter dekanen anfører som dokumentation for min groft forsømmelige adfærd

"Som dokumentation for den groft forsømmelige adfærd kan nævnes:" (ifølge dekanen):

1. " ... Blandt andet har Nyborg i flere år angivet, at analysen af voksne var baseret på 52 personer og ikke 62."

Kommentar: Præliminære analyser blev foretaget på 52 forsøgspersoner (fp.) med komplette data. ISIR2001 analysen blev gennemført efter opdatering, således at der nu var 62 fp. hvoraf nogle dog ikke havde komplette data. På grund af en skrivefejl anførtes stadig 52 fp.. Den endelige publicerede 2005 analyse indeholdt 62 fulde datasæt. Variationerne i fp. og data påvirkede kun resultatet på anden eller tredje decimal, dvs. gjorde ingen forskel i det overordnede resultat.

2. "... påstande, der ikke er korrekte."

a) "Data on children participating in the cross-sectional parts of the study were included in the present study. Dette er ikke korrekt, idet børn fra den første tværsnitsundersøgelse i 1976 overhovedet ikke bidrager med data ..."

Kommentar: Det er korrekt at børn fra den første tværsnitsundersøgelse ikke bidrog, men det gjorde børn fra projektets anden tværsnitsundersøgelse.

b) "Imidlertid afslører den udbredte anvendelse af repeated measures i den longitudinale undersøgelse, at der i strid med det erklærede også indgår data, som stammer fra gentagne undersøgelser af de samme børn."

Kommentar: Jeg anfører i den endelige 2005 artikel at der indgår fp. fra både longitudinelle og tværsktionelle faser. Det er meningsforstyrrende, når det sagkyndige udvalg sammenblander præliminære og kursorisk omtalte tidlige analyser, hvor observationer fra gentagne undersøgelser bruges (og som Institutleder Jens Mammen tjenstligt beordrede mig til at offentliggøre på min private hjemmeside under protest), med den publicerede 2005 artikel hvor kun data fra enkeltundersøgelser optræder.

c) "The particular selection procedure resulted in a total of 376 children and adults. Dette er også ukorrekt ..."

Kommentar: Ja, dette er ukorrekt, men i det øjeblik "children and adults" erstattes med "observations", er udsagnet korrekt. Der er tale om en skrivefejl. I øvrigt har denne analyse

intet at gøre med den endelige analyse præsenteret i 2005 artiklen. Det er forvirrende at det sagkyndige udvalg vedblivende henviser til præliminære analyser, og ikke koncentrerer sig alene om de data der anvendes i den endelige afrapportering.

3. a) " ... gentagne anvendelser af flere forkerte formler ... "

Kommentar: Den internationale sagkundskab har forskellige meninger om hvilke formler der bør anvendes. En diskussion heraf bør forbeholdes det internationale forskerforum, og ikke anvendes snævert i en lokal disciplinærsag.

b) *Endvidere mener udvalget ikke, at Nyborg på noget tidspunkt i tilstrækkelig grad har redegjort for undersøgelsens design.*

Kommentar: Der er tale om faglig uenighed her. De anonyme review-eksperter fastslog uden tøven, at jeg rent faktisk har redegjort i tilstrækkelig grad for designet (se bilag 13). I øvrigt har forlaget begrænset indleverede artikler til max. 5.000 ord (inkl. tabeller og figurer) og jeg fik førsteudkastet til 2005 artiklen tilbage, ledsaget af et ufravigeligt krav om væsentlige reduktioner. Specialister indenfor feltet finder, med andre ord, at det sagkyndige udvalg tager fejl her.

4. " ... care must be taken to ensure that the dropouts does not introduce bias. We have not registered any indication that Nyborg is aware of this basic problem, and the reader is never alerted to the mixed and unsatisfactory composition of the data set "

Kommentarer: Her tager udvalget fejl på flere punkter. For det første omtales drop-out problemet på side 208 i min 2003 bog. Dernæst gøres læseren af 2005 artiklen opmærksom på, at der er tale om et "cohort-sequential study", dvs. at der er tale om et mikset tværsektionelt-longitudinelt design. Endeligt anføres det, at "No particular pattern of reasons for refusing to participate could be spotted in retrospect". Endeligt ekspliciteres det, at det tværsektionelle 1976 studie indeholder "... only some cognitive testing ...", og at det longitudinale studie inkluderer "... repeatedly tested groups ..." (p. 500).

I øvrigt kunne det sagkyndige udvalg selv (ud fra den dem tilgængelige datadisk: CD: bilag 26) have kontrolleret hvorvidt drop-outs havde introduceret en bias i undersøgelsen. Man kunne have sammenlignet scores fra de fp. der var undersøgt gentagne gange over en årrække med de fp. der kun var undersøgt en gang.

En sådan analyse har jeg gennemført, og den viser at eksklusion af personer, der er undersøgt flere gange, ikke ændrer hovedkonklusionen.

Datasæt: "Adult complete data set N = 62 with ID and sex.sta". Analyse: Selektionskriterium: SUBJECT2 >= 80000; Point-biserial beregnet vha. formlen $rpbs = d / (2 * \sqrt{KVROD((d^2)/4 + 1)})$; efter Jensen, 1998); rpbs'erne blev sat ind i korrelations-matrisen, og den resulterende file blev som sædvanligt kørt gennem Statistica hierarkisk factor-analyse program med indstillingen max. 7 faktorer. Resultat: rpbs = 0,334.)

Noter for ikke-specialister:

En punkt-biserial korrelation er simpelthen en Pearson product-moment korrelation der udtrykker relationen mellem en metrisk variabel (f.eks. test scores) og en dikotom variabel (her køn, hvor konventionelt mænd = 1 og kvinder =0).

Det, at man sætter rpbs ind for hver sub-test i korrelationsmatricen for faktoranalyse, har ingen effekt på faktorstrukturen og har ingen praktisk betydning for

de enkelte subtests g -loadings (kongruenskoeficient = 0,999 for alle batterier: Jensen, 1998, p. 542).

Det er værd i forbifarten at notere sig, at dette sidste står i åbenbar kontrast til ph.d. stipendiat Pia Ankersens m.fl.s påstand i Jyllandsposten (se nedenfor) om at jeg, ved at bruge denne fremgangsmåde, har "fabrikeret" en kønsforskel der ikke findes i mine data.

Ikke desto mindre optræder Ankersens dagspresseindlæg som et vigtigt indlæg i sagen mod mig. Således anfører tidligere dekan Tom Latrup-Pedersen i brev af 27. januar 2006 til mig, at når han brugte følgende passus: "When members of the scientific community raise serious and articulated allegations questioning the empirical validity of research findings, and/or the ethics of procedure, ...", så tænkte han på "... Pia Ankersens klage den 30. september 2003 til institutleder Jens Mammen og sammes klage den 5. oktober 2003 til Aarhus Universitets Praksisudvalg. Endvidere tænkte jeg på Pia Ankersens, Jørgen Poulsens og Siggie Kristoffersens kronik den 9. oktober 2003 i Jyllandsposten (med teknisk tillæg på nettet), ...".

Jeg finder det foruroligende at en dekan for en væsentlig dels vedkommende synes at basere universitets kritik af min kønsforskning, på en aviskronik med så graverende fejl.

Hovedpointen med at kommentere drop-out problemet her er imidlertid, at køn loader på g uanset om fp. er undersøgt mange gange eller kun en gang. Den oprindelige g -loading på 0,28 for totalpopulationen og g -loadingen på 0,334 for engangsundersøgte fp. ligger begge indenfor det udfaldsområde af resultater som det sagkyndige udvalg selv finder under anvendelse af mange forskellige faktoranalytiske designs. Med andre ord er konklusionen den, at drop-outs ikke kan antages at have introduceret en uacceptabel bias i undersøgelsen, og at læserne af den endelige 2005 rapport derfor ikke, som påstået, er blevet foreholdt vigtig viden om afgørende begrænsninger i undersøgelsen.

5. " ... Nyborgs gentagne forsikringer om, at dataindsamlingen stadig foregår ... " bidrager til det indtryk " ... at studiets design er forvirrende og mangelfuldt ... ".

Kommentar: Udvalget tager også fejl her. At studiets design forekommer forvirrende og mangelfuldt skyldes i alt væsentligt at jeg tjenstligt blevet afkrævet en detaljeret redegørelse for et 30årigt kohort-sekventielt kæmpeprojekt med 1.200+ variabler på et tidspunkt, hvor hele databasen var under opbygning, og hvor jeg ud af dette kæmpemateriale havde valgt blot at definere og analysere 20 snævert udvalgte variabler for et begrænset antal personer. På et møde med dekan og institutleder blev jeg ikke desto mindre anmodet om ud fra hukommelsen, og uden en forudgående skriftlig orientering om mødets dagsorden, at redegøre nøje for præcist hvor mange børn der indgik i hver enkelt fase helt tilbage til 1981. På en sådan baggrund er det let at give det indtryk at forvirring og mangelfuldhed er fremherskende. Imidlertid burde dekanen, efter min opfattelse, i sin indberetning have anført, at der siden hen rent faktisk er fremsendt klare grafiske og numeriske oversigter over hele designet, med nøje angivelse af visiteringstidspunkter og antal fp.

Endvidere har udvalget, så vidt jeg kan skønne, ingen basis for at fastslå, at der ikke til stadighed indsamles data i projektet. I den udstrækning finansiering tillader det søges voksendata fortsat komplementeret. Hertil kommer, at mange rådata for alle faser af projektet endnu ikke er lagt på elektronisk form.

6. *Udvalget efterlyser "... en diskussion af den omfattende anvendelse af "mean substitution" ..."*

Kommentar: Her blander udvalget igen de præliminære metodeorienterede analyser sammen analyser af det datamateriale der anvendtes til at dokumentere den observerede kønsforskel, i 2005 artiklen. I denne endelige afrapportering har "mean substitution" og problemet med $N=325$ ingen som helst relevans.

7. *"Nyborgs vedholdende nægtelse af andre forskeres adgang til sine data ..."*

Kommentar: Intet kunne være mere forkert. Jeg har aldrig nægtet nogen adgang til data, og vil nu gerne have denne fejlagtige påstand dokumenteret! Sagen er, at "Politiken" fik mine data allerede i januar 2002, men mistolkede dem og "lavede" en helt urealistisk kønsforskel på 27%, til trods for at 27% hverken optrådte i de to artikeludkast jeg fik tilsendt eller i databladene. Jeg indtog derefter den holdning (skriftligt), at data fra de komplicerede analyser af kønsforskelle fremover kun ville blive udleveret sammen med en nøjere beskrivelse af metoder og fund. Dette synspunkt bifaldt daværende rektor Niels Sidenius. Han medgav skriftligt, at en forsker må gives rimelig tid til endelig afrapportering). Rapporten publiceredes som bekendt i 2005.

8. *"... yderlige indikation af den manglende dokumentation og manglende omhu ..." er "... at man ikke har været i stand til at reproducere Nyborgs resultater præcist ..."*

Kommentar: Dette er i bedste fald misvisende. Det er alment kendt at i faktor analyse er antallet af permutationsmuligheder meget stort, og at forskellige modeller giver kvantitativt forskellige resultater, der dog kvalitativt ligger indenfor samme forventede udfaldsspektrum. Dette har intet at gøre med manglende dokumentation og da slet ikke med manglende omhu. Det sagkyndige udvalgs egne beregninger dokumenterer også netop dette spektrums relevans ved at anføre, at udvalget ud fra mine data kvalitativt opnår de samme resultater som Nyborg. Rent faktisk giver en af udvalgets mange analyser en loading på 0.277, hvilket kommer forførelserisk tæt på min publicerede loading på 0.274 (p. 502). Pudsigt nok ligger gennemsnittet af udvalgets mange analyser, som noteret af Rushton (se bilag 10), på 0.27. (se i øvrigt også kommentarerne til punkt 9 nedenfor).

9. *"One cannot tell if the differences in the means of the 20 test variables are due to a difference in the g-factor or due to a difference among the primary factors."*

Kommentar: Hovedformålet med undersøgelsen var at se om køn loadede på g -faktoren. Det er derfor yderst beklageligt, at det sagkyndige udvalg helt undlader at fortælle, at jeg har lavet en korreleret vektorberegning (tabel 2 i 2005 artiklen, p. 503). Denne form for analyse har netop har til formål at undersøge hvorvidt kønsforskellen ligger i g .

Note for ikke-specialister

Metoden er beskrevet i detaljer i Appendix B i Jensen (1998, pp. 589-591). Kort fortalt går metoden ud på at rang-orden korrelere hver enkelt sub-tests g -loading (korrigeret for reliabilitet, dvs. efter at g -loadingen er divideret med kvadratroden af hver sub-tests reliabilitetskoefficient) med kønsforskellen i de enkelte sub-tests (udtrykt ved d -effekter, som også korrigeres for reliabilitet). Ideen er, at kønsforskelle ligger i g i den udstrækning kønsforskelle i d øges med stigende g -loading af subtests.

Den korrelerede vektorberegning viste en Spearman rank-order r_s (one-tailed) = 0,412; $p = 0,035$, hvilket bekræfter at kønsforskellen loader signifikant på g . Dette udelukker

naturligvis ikke at der kan være kønsforskelle på gruppefaktor niveau, og at nogle af disse forskelle rent faktisk kan påvises at være i kvinders favør.

Ved at undlade at nævne vektorberegningen og dens udfald i sin rapport gør det sagkyndige udvalg sig, efter min opfattelse, på dette specifikke punkt skyldig i direkte misinformation i en sag om kønsforskelle i *g*, der ifølge udvalgets kommissorium kunne få (og fik) disciplinære konsekvenser.

10. "At Nyborg enten ikke har forstået faktoranalysens begrænsninger eller ignorerer dem og undlader at komme ind på dem ...".

Kommentarer:

a) Jeg hverken undlader eller ignorerer begrænsningerne. Således anfører jeg rent faktisk (p. 501 i 2005 artiklen), at den hierarkiske faktoranalyse viser "...little dimensional contamination ...", hvilket begrundes at jeg anvender netop denne analyseform.

b) Endvidere anfører udvalget "...that since the hierarchical factor analysis involves a number of choices it is good scientific practice to discuss robustness of the results under various choices."

Kommentar: Det er naturligvis absolut sandt, at man bør diskutere resultatets robusthed under forskellige analysebetingelser - hvor dette er påkrævet! Imidlertid er et sådant krav helt urealistisk, givet at *Personality and Individual Differences* (og mange andre internationalt respektable tidsskrifter) har meget eksplicite pladsbegrænsninger (se kommentar til punkt 3).

I øvrigt er der bred enighed i psykometri-kredse om, at den af det sagkyndige udvalg krævede diskussion overhovedet ikke er påkrævet som led i "... good scientific practice ..." i forbindelse med bestemmelse af *g*. Grunden hertil skal findes i følgende empiriske forskning, der klart viser, at de forskellige faktoranalysemetoder estimerede "true *g*" så tæt, at "...there was hardly any basis for choosing between them" (e.g. Jensen & Weng, 1994).

Jensen and Weng (ibid.) thus raised the question of how invariant *g* is across various methods of factor analysis. They used six different factor analytic methods on four simulated data matrices where the factors were exactly known. They also used nine different factor analytic methods on a real correlation matrix with twenty-four tests taken by 145 grade 7 and 8 students. The average congruence coefficients between the true *g* factor and the *g* factors derived from the various methods amounted to + 0.998 (range + 0.997 - + 0.999). This applied even if some of the artificial matrices were deliberately designed to "trick" deviating estimates. For the real data, the forty-five congruence coefficients between the ten *g* vectors ranged from + 0.991 to 1,000 (average + 0.995). They concluded that all the different methods of factor analysis estimated the true *g* so closely that **there was hardly any basis for choosing between them** (my emphasis).

Moreover, Ree and Earles (1991) factor analyzed data for 9,173 recruits taking the Armed Services Vocational Aptitude Battery (ASVAB) with 14 different methods and derived 14 different *g* factor scores for each recruit. The many different methods resulted in very little variation among the obtained *g* factors (average correlation +.984).

Senest fastslår Jensen (1998, p. 83) at, "... whatever variation exists among the myriad estimates of *g* that have been reported since the beginning of factor analysis, **exceedingly little of it can be attributed to differences in the methods of factor analysis employed.**" (my emphasis).

Læg hertil, at det sagkyndige udvalgs mange analyser gav resultater der konfirmerede ikke blot dette, men også min hovedkonklusion (se kommentarer til punkt 7).

Det er derfor helt uforståeligt at det sagkyndige udvalg afkræver mig en omfattende diskussion af ting, der for længst er empirisk påvist at være irrelevante, og som peer-review eksperterne ved den ansete internationale journal jeg publicerede resultatet i, ikke anså for nødvendig.

Ad Loyalitetsforpligtelse

Dekanen har endnu ikke givet mig en længe lovet oversigt over de oplysninger jeg har givet om sagen til internationale kolleger og som universitetet tolker som illoyal adfærd. Derfor anmoder jeg fortsat om at få tilsendt denne oversigt - også efter at jeg er fritaget for tjeneste.

Det står fast, at det var ukorrekt af mig at anføre, at praksisudvalget anklagede mig for præmatur publikation; det var Institutleder Jens Mammen der var ansvarlig for denne anklage. Om denne fejl stiller universitetet i et urimeligt dårligt lys er et interessant spørgsmål.


Afsluttende bemærkninger.

Det er mit håb at ovenstående kommentarer mere end antyder, at jeg ikke har udvist "... en groft uagtksom adfærd ...", at jeg ikke "... på utilbørlig vis har vildledt om egen videnskabelig indsats og resultater..." og at min undersøgelse af kønsforskelle i almen intelligens, g, ikke "... mangler enhver troværdighed og i realiteten er uden værdi."

Jeg tolker også de mange udenlandske eksperters vurderinger, samt udtalelserne fra adskillige tidligere og nuværende præsidenter for relevante faglige fora, og fra de tre redaktører af "Personality and Individual Differences" (bilag 10 - 25) derhen, at jeg næppe kan siges at have gjort mig skyldig i grov forsømmelighed således at der er grundlag for at indlede en disciplinærundersøgelse mod mig for tjenstlig forseelse.

Jeg beklager at dekanen, "... under henvisning til den store offentlige interesse for og polemik om "Skanderborgprojektets" resultater ..." har valgt at offentliggøre udvalgets rapport og kommissorium på fakultetets hjemmeside, uden dog at vedhæfte mine kommentarer til udvalgets vurdering, med det resultat at dele af pressen på forhånd har dømt mig uredelig.

Med venlig hilsen


Helmuth Nyborg
Adslev Skovvej 2
8362 Hørning