

„Nordisk Psykologi“s  
monografiserie nr. 7

# ROSENZWEIGS BILLED-FRUSTRASJONSTEST

EN DISKUSJON AV DENS ANVENDBARHET  
BELYST VED ANDRES OG EGNE UNDERSØKELSER

BJØRN CHRISTIANSEN

Institutt for samfunnsforskning, Oslo

I KOMMISJON HOS  
EYNAR MUNKSGAARDS FORLAG, KØBENHAVN  
A/S BOKHJØRNET, OSLO

Trykt med bidrag fra  
Norges almenvitenskapelige forskningsråd



1955

## Forord

Bruken av flere tester med spesielle funksjoner til foretrengsel for én test av mer generell karakter, er en av de mest påfallende utviklingslinjer i moderne psykodiagnostikk. Denne utvikling blir for øvrig fulgt av vedvarende forsøk på å gjøre de projektive personlighetsprøvers grunnlag mest mulig eksplisitt og deres anvendelse mest mulig vitenskapelig forankret.

Det er for tiden blant norske psykologer en meget stor interesse for personlighetspsykologi. I særlig grad synes interessen å være konsentrert om projektive diagnostiske metoder. Ennå er ikke noen slik metode blitt systematisk utprøvet under norske forhold, og behovet for vitenskapelige undersøkelser og drøftinger synes i høy grad å være til stede.

Det foreliggende arbeid tar sikte på å gi en bred fremstilling av Rosenzweigs *billed-frustrasjonstest*. Dette er en prøve med en relativt spesiell funksjon. Dens oppgave er utelukkende å belyse enkelte sider ved et individs personlighet. Disse sider er riktignok av sentral betydning, men prøven vil alltid måtte kompletteres av andre prøver hvis man tar sikte på å gi en mer omfattende personlighetsdiagnose. Hensikten med arbeidet er i første rekke å gi en vurdering av testen basert på resultatene av empiriske undersøkelser. Ved siden av fyldige beskrivelser av andres forskningsresultater vil det også bli presentert data fra egne undersøkelser.

Testens mer spesielle psykometriske sider vil bli grundig drøftet, men det vil også bli lagt stor vekt på å trekke frem endel teoretiske spørsmål som knytter seg til projektive prøver mer generelt. Fremstillingen vil derfor på enkelte punkter også kunne være av interesse for psykologer som ikke er spesielt innstilt på selv å ta billed-frustrasjonstesten i bruk.

Til gjennomføringen av mine egne undersøkelser har jeg hatt hjelp av flere. Stud. psychol. Aud Myhr hjalp meg med innsamlingen av empirisk materiale, cand. psychol. Kjell Larsson og cand. psychol. Arne Lie med skåringen og tabuleringen av materialet, og cand. psychol. Arvid Amundsen med den senere statistiske bearbeiding. Uten deres villighet til å gjøre mye arbeid for liten lønn ville undersøkelsen ikke vært mulig.

Universitetsstipendiat Arvid Ås, dosent Ragnar Rommetveit, stipendiat Sverre Brun-Gulbrandsen, professor Burton R. Fisher og sekretær Eivind Baade har lest igjennom manuskriptet under utarbeidelsen og fremkommet med råd og vink av både faglig og fremstillingsmessig art.

Jeg takker alle for hjelpen og ikke minst for den velvillighet og interesse de har vist arbeidet. Til slutt vil jeg også rette en takk til Norges almenvitenskapelige forskningsråd for støtte til trykning av arbeidet.

Oslo i oktober 1954.

*Bjørn Christiansen.*

## INNHOOLD

| Kapitel  | Side |
|--|------|
| I. INNLEDNING .....  | 7    |
| II. TESTENS HISTORISKE BAKGRUNN .....  | 9    |
| III. BESKRIVELSE AV TESTEN .....   | 16   |
| IV. TESTENS STATISTISKE NORMER .....   | 22   |
| V. TESTENS RELIABILITET .....  | 26   |
| Skåringsreliabilitet .....   | 26   |
| Retest-reliabilitet (stabilitet) .....   | 28   |
| Indre reliabilitet (konsistens) .....  | 31   |
| VI. TESTENS VALIDITET: 1. GENERELLE BETRAKTNINGER..  | 35   |
| VII. TESTENS VALIDITET: 2. DIFFERENSIERINGSEVNE .....  | 41   |
| Differensieringsevne overfor forskjellige sosiale grupper....  | 41   |
| Differensieringsevne overfor grupper med forskjellig klinisk<br>diagnose .....   | 45   |
| Differensieringsevne overfor grupper med forskjellig frustra-<br>sjonstoleranse .....  | 48   |
| Differensieringsevne overfor grupper med forskjellig sosial<br>status .....  | 52   |
| VIII. TESTENS VALIDITET: 3. SAMSVAR MED UAVHENGIGE<br>KRITERIER .....  | 54   |
| Samsvar med andre tester .....   | 54   |
| Samsvar med manifest atferd .....  | 59   |
| Eksperimentelle undersøkelser .....  | 64   |
| IX. TESTENS VALIDITET: 4. NIVÅFORANKRING .....   | 70   |
| Et historisk tilbakeblikk .....  | 70   |
| Nyere undersøkelser .....  | 72   |
| Forholdet mellom atferdsnivåer .....   | 74   |
| Nivåbestemmende forhold .....  | 77   |
| Foreløpig konklusjon .....   | 81   |
| X. EGNE EMPIRISKE BIDRAG .....   | 82   |
| Bakgrunn for egne undersøkelser .....  | 83   |
| Undersøkelse av skåringsreliabiliteten .....   | 84   |
| Undersøkelse av indre konsistens .....   | 87   |
| Sammenligning med amerikanske kategorinormer .....   | 92   |
| Sammenligning med skolehjemsgutters reaksjonstendenser   | 96   |
| Undersøkelse av korrelasjonen med tilpassingsvurderinger ..  | 97   |
| Undersøkelse av testens nivåforankring .....   | 100  |
| Undersøkelse av testens validitet .....  | 102  |
| Undersøkelse av korrelasjonen med en innstillingsskala .....   | 106  |
| Sammendrag og konklusjoner .....   | 114  |
| LITTERATUR .....   | 116  |
| APPENDIKS I: Norsk revisjon av billed-frustrasjonstesten for<br>voksne — utarbeidet og brukt i forbindelse<br>med egne undersøkelser ..... | 123  |
| APPENDIKS II: Norsk revisjon av skåreskjema til billed-frustra-<br>sjonstesten — utarbeidet i forbindelse med egne<br>undersøkelser .....  | 131  |



## Kapitel I

### Innledning

Bruken av og interessen for projektive prøver blir stadig mer omfattende. Nye prøver blir konstruert og gamle revidert. I dag finnes neppe en klinisk psykolog med respekt for seg selv og sitt yrke som ikke betjener seg av projektive prøver som et personlighetsdiagnostisk instrument.

Også innenfor hva som tradisjonelt betegnes som sosialpsykologi, er projektive prøver i ferd med å vinne innpass, dels som et middel til forklaring av sosial atferd, og dels som et utgangspunkt for sosialterapeutiske inngrep (action research).

En amerikansk projektiv prøve av relativt ny dato er Saul Rosenzweigs *billed-frustrasjonstest*. Denne prøve har fått en ualminnelig hurtig utbredelse i de aller siste år, og blant projektive personlighetsprøver i U.S.A. er det faktisk — ifølge Buros' *Fourth Mental Measurements Yearbook* fra 1953 — bare *Rorschach* og *Thematic Apperception Test (TAT)* som kan fremvise et større antall enkeltpublikasjoner. Heller ikke i U.S.A. foreligger det ennå noen mer omfattende behandling av denne prøve, og den interesserte psykolog, psykiater eller sosiolog er derfor utelukkende henvist til et utall begrensede og ofte svært spesielle artikler i bøker og tidsskrifter.

Betegnelsen projeksjonstest eller projektiv teknikk er i dag innarbeidet og hyppig brukt innenfor så vel personlighets- som sosialpsykologi. Termen er av ganske ny dato. Den ble først introdusert og brukt på en konsistent måte av Lawrence K. Frank i 1939. Samme år ble det publisert et av de første større eksperimentelle arbeider med forankring i psykoanalytiske begreper. En sikter til Dollard, Doob, Miller, Mowrer og Sears' bok: *Frustration and Aggression* (15). I en større tidsskriftartikkel året før blir frustrasjon som eksperimentelt problem tatt opp til drøfting på bredt grunnlag (60). Og et par år senere følger Barker, Dembo og Lewins berømte eksperimentelle undersøkelse av frustrasjon og regresjon (4).

Rosenzweigs billed-frustrasjonstest eller *B-F-testen*, som vi har valt å kalle den, representerer på sett og vis en syntese av de ovenfornevnte utviklingslinjer. På den ene side er den konstruert som en

prosjektiv prøve. På den annen side er den *orientert* i retning av frustrasjon og reaksjoner på frustrasjon.

Rosenzweigs B-F-test foreligger i dag i to utgaver, én utgave beregnet spesielt for voksne og én for barn. Utgavene er i prinsippet svært like. Utgaven for voksne er blitt oversatt til norsk for noen år siden av en gruppe psykologistudenter. Utgaven for barn er av nyere dato, og det empiriske materiale som foreligger om denne utgave, er mer begrenset. Også denne utgave er i det aller siste blitt oversatt til norsk i forbindelse med et test-psykologisk forskningsprosjekt ved Dr. Nic Waal's Institutt i Oslo.

Det er på det nåværende tidspunkt ikke mulig å gi noen endelig vurdering av B-F-testens anvendbarhet. Tiden den har vært i bruk, er relativt kort. En rekke undersøkelser over testen er imidlertid blitt utført — særlig i U.S.A. — i de siste år, slik at det i dag til en viss grad er mulig å underbygge vurderinger med empirisk materiale.

Vi vil i det foreliggende arbeid først og fremst konsentrere oss om den utgave som er beregnet for voksne. Det vil først bli gitt en kort historikk. Dernest vil prøvens reliabilitet og validitet bli tatt opp til drøfting. Dette vil utgjøre arbeidets ene hovedformål.

Det annet formål med arbeidet vil være drøfting og analyse av en del data basert på egne undersøkelser med den norske revisjon av testen.

I forbindelse med den systematiske fremstilling av andres undersøkelser vil det bli gitt relativt fyldige litteraturhenvisninger. Ved hjelp av disse vil interesserte lett kunne finne frem til spesielt bakgrunnsmateriale, standarder og normer. Henvisninger vil utelukkende bli gitt til arbeider over testen som foreligger i Norge.

Litteraturlisten bak representerer ikke noen komplett bibliografiliste over testen. En liste over samtlige publiserte arbeider til og med 1951 vil dem interesserte finne i *The Fourth Mental Measurements Yearbook* (11). Referanselisten her omfatter i alt 77 arbeider. Av disse er 25 magister- eller doktorgradsavhandlinger, 4 er henvisninger til lærebøker hvor testen er blitt gitt en spesiell behandling, 5 er til abstracts, og resten — 47 — er til tidsskriftartikler.



## Kapitel II

### Testens historiske bakgrunn

Et virkelig grundig kjennskap til B-F-testen kan man vanskelig oppnå uten gjennom en viss innsikt i testens historiske bakgrunn. Testens forutsetninger er på det nærmeste knyttet sammen med Rosenzweigs problemstillinger og forskning allerede fra begynnelsen av 30-årene.

Det som i særlig grad sto i brennpunktet for Rosenzweigs interesse på det tidspunkt, var *fortrengning* som psykologisk problem. I en serie eksperimenter ble forholdet mellom gjentakingspreferanser og suksess og nederlag ved oppgaveløsning tatt opp til undersøkelse (53).

Resultatene han kom fram til, viste stor overensstemmelse med det psykoanalytiske lyst- og realitetsprinsipp. Yngre barn viste en uttalt tilbøyelighet til å gjenta oppgaver som de tidligere hadde måttet å løse, mens noe eldre barn særlig la for dagen gjentakingspreferanser for oppgaver de ikke hadde greidd.

I en annen serie med undersøkelser var utgangspunktet å finne frem til forholdet mellom oppgaveløsning og oppgavehukommelse (55). Det viste seg at enkelte barn hadde en tendens til fortrinnsvis å huske oppgaver de hadde greidd å løse, andre fortrinnsvis oppgaver de ikke hadde greidd, mens det hos atter andre ikke var noen uttalt preferanse hverken for løste eller uløste oppgaver. Fordelingen syntes å være normal. En sammenligning av barna i de to ekstremgrupper viste imidlertid at barna som fortrinnsvis husket løste oppgaver, stort sett var mer mentalt modne og også hadde en mer utviklet stolthetsfølelse — ifølge lærervurderinger — enn de som fortrinnsvis husket uløste oppgaver.

Barna som ble brukt som forsøkspersoner (*fpp.*) i dette eksperiment, var de samme som var nyttet tidligere. En nærmere analyse brakte for dagen at det var de barna som fortrinnsvis husket løste oppgaver, som hadde gjentakingspreferanser for uløste, og på den annen side — de barna som fortrinnsvis husket uløste oppgaver — som hadde gjentakingspreferanser for løste.

Resultatene antyder at det hos de mentalt eldste barna fant sted en fortrenning, en hukommelsessvikt, i forhold til uløste oppgaver.

De mentalt eldste barna var også preget av en sterkere stolthetsfølelse. Det er rimelig å anta at medmindre et nederlag blir oppfattet som et sår på selvrespekten, vil det ikke utløse behov for sosial rettferdiggjørelse eller fortrenkning. Eksperimentet antyder med andre ord at *en viss type* fortrenkning først blir tatt i bruk relativt sent i et barns utvikling — når barnets selvfølelse og stolthetsfølelse har nådd et bestemt nivå.

Nå er det naturligvis ikke slik at stolthet er en faktor som alle eldre barn har like mye av eller som gir seg uttrykk i like sterk grad i alle situasjoner. Oftest er det slik at enkelte situasjoner appellerer mer til stolthetsfølelse enn andre. Også hos voksne er det rimelig å anta at dette er tilfelle.

For å prøve denne antagelse mer direkte foretok Rosenzweig en tredje type av eksperimenter. Også nå var hensikten å finne frem til hvordan hukommelse var avhengig av suksess eller nederlag ved oppgaveløsning (54, 61).

Eksperimentet ble utført på to grupper voksne. Begge fikk seg forelagt en serie puslespill som de ble satt til å løse. For den ene gruppen ble puslespillene presentert på en helt uformell måte, og fpp. ble motivert ut fra ønske om å hjelpe til å klassifisere puslespillene for senere bruk. For den andre gruppen ble puslespillene introdusert som en intelligenstest. I begge grupper ble fpp. plutselig avbrutt midt i spillene. Spillene ble så samlet sammen og fpp. anmodet om å notere hvilke spill de hadde forsøkt seg på.

Resultatet av eksperimentet var at i gruppen som arbeidet under uformelle betingelser, var det stor overvekt av dem som best husket uferdige oppgaver, mens forholdet i den gruppen som arbeidet under intelligenstest-betingelser, var stikk motsatt. Her husket flertallet fortrinnsvis de oppgaver de var blitt ferdige med.

Resultatet støtter den tidligere antagelse at en viss stolthetskomponent må være til stede hvis fortrenkning skal finne sted.

En side ved resultatene som Rosenzweig la betydelig vekt på, var de individuelle forskjeller som fremkom. Enkelte fpp. som ble prøvd under uformelle betingelser, la for dagen større grad av fortrenkning enn enkelte av dem som tok oppgavene som intelligenstest.

Som forklaring på forholdet antok Rosenzweig at han i disse tilfelle sto overfor bestemte utslagsgivende personlighetstrekk. Enkelte fpp. hadde liksom en disposisjon til å reagere med fortrenkning av nederlags-situasjoner, andre en uttalt motstand mot en slik reaksjonsform.

Mer eller mindre tilfeldig hadde Rosenzweig i løpet av eksperimenteringen lagt merke til at fpp. han arbeidet med, hadde forskjellige tilbøyeligheter når det gjaldt umiddelbar reaksjon på frustrasjon eller nederlag. Enkelte hadde en tendens til å bebreide seg selv eller omgivelsene, andre en tendens til å dekke over, bortforklare eller finne unnskyldende momenter, etc. Så interessant ble Rosenzweig i disse forskjellige reaksjonstyper at han gikk i gang med å systematisere sine inntrykk og konstruere et system til klassifikasjon av såkalte «apperceptive typer av bevisste reaksjoner på frustrasjon» (56).

Han gjorde seg til talsmann for en inndeling mellom *ekstrapunitiv*, *intropunitiv* og *impunitiv* reaksjonstyper — og hevdet at disse tre typer hadde sine bestemte så vel emosjonelle som vurderingsmessige komponenter.

## SKJEMA 1

### *Rosenzweigs opprinnelige klassifiseringsskjema*

| Komponenter | Ekstrapunitiv   | Intropunitiv  | Impunitiv   |
|-------------|---|---|---|
| Emosjoner   | Sinne og vrede  | Ydmykelse og skyldfølelse   | Forlegenhet og skamfølelse                                  |
| Vurderinger | Bebreider andre:<br>„Du har grist meg til. Jeg skal sørge for at du får det igjen.“ | Bebreider seg selv:<br>„Hvordan kunne jeg gjøre noe slikt. Jeg kan aldri tilgi meg selv.“ | Tilgir:<br>„Det kunne ikke unngås. La det hele være glemt.“ |

Det første spørsmål han stilte seg — med bakgrunn i klassifiseringsskjemaet, var hvorledes de forskjellige reaksjonstyper på frustrasjon sto i forhold til hukommelse. Han antok på grunnlag av de foreløpige og tilfeldige observasjoner han hadde gjort, at personer som hadde en tendens til å huske nederlag, var slike som hadde for vane umiddelbart å bebreide seg selv eller andre når de opplevde frustrasjon, mens de som hadde en tendens til å huske seirer, var slike som pleide å dekke over — reagere impunitivt — overfor frustrasjon.

Forklaringen mente han å finne i at både den ekstrapunitive og intropunitive reaksjonstype — i motsetning til den impunitive — omfattet en aggressiv komponent. Denne komponent antok han det var som virket fremmede på en vedvarende hukommelse av den frustrerende situasjon.

Det neste problem som opptok Rosenzweig, var å finne frem til en metode eller et instrument som på en objektiv måte kunne registrere de forskjellige reaksjonstendenser slik at han kunne bli i stand til å prøve sine antagelser (67).

En av de første fremgangsmåter han konsentrerte seg om, var å gi en serie oppgaver av dels språklig, dels aritmetisk type. Oppgavene ble presentert én om gangen. Løsningstiden var begrenset og meget kort. Halvparten av oppgavene var meget vanskelige, halvparten faktisk uløselige. Fpp. ble instruert om å løse de oppgavene de fikk seg forelagt, og hvis de ikke fikk det til, angi om de anså oppgavene som løselige eller uløselige.

Opgavene var slik konstruert at alle fpp. opplevde flere skuffelser eller nederlag. En tilbøyelighet til å oppfatte oppgaver som løselige ble tolket som en ikke-ekstrapunitiv tendens, mens påstander om uløselighet ble sett på som uttrykk for ekstrapunitive tendenser.

Fremgangsmåten hadde en innlysende ulempe, den ga ikke noe grunnlag for differensiering mellom intropunitive og impunitive reaksjonstendenser.

I *Exploration of Personality*, utgitt i 1937, har Rosenzweig en kort artikkel, hvor han redegjør for metoden (59). I samme artikkel forsøker han også å sette sitt tidligere klassifiseringsskjema inn i en bredere psykoanalytisk sammenheng.

## SKJEMA 2

### *Rosenzweigs reaksjonstyper med hypotetisk tilknyttede psykoanalytiske begreper*

| Type-navn                     |                             | Ekstrapunitiv                   | Intropunitiv                       | Impunitiv                        |
|-------------------------------|-----------------------------|---------------------------------|------------------------------------|----------------------------------|
| Pathognomonisk bilde          | Emosjoner:                  | Sinne (med fiendtlighet)        | Skyldfølelse (med anger)           | Forlegenhet (med skamfølelse)    |
|                               | Vurderinger:                | Fordømmelse av ytterverden      | Fordømmelse av eget selv           | Tilgivelse av andre og eget selv |
| Dynamisk og genetisk grunnlag | Instinkter:                 | Aggressive behov rettet utover  | Aggressive behov rettet innover    | Erotiske behov                   |
|                               | Identifiseringsinnstilling: | Frykt for straff                | Behov for straff                   | Frykt for å miste kjærlighet     |
| Psykopatologiske aspekter     | Forsvarsholdning:           | Projeksjon                      | Forskyvning og isolasjon           | Fortrengning (med selvbedrag)    |
|                               | Mentale forstyrrelser:      | Paranoid (paranoid schizofreni) | Tvangsmessig (kataton schizofreni) | Hysterisk (hebefren schizofreni) |
|                               | Libidinøse typer (Freud):   | Narsistisk                      | Kompulsiv                          | Erotisk                          |

I sitt fortsatte arbeid med hukommelses- og glemselsprosesser utvider han etter hvert sin problemstilling til også å omfatte hypnotiserbarhet, idet han antar at det ikke bare foreligger korrelasjon mellom impunitive reaksjoner på frustrasjon og fortrengningstendenser, men også korrelasjon mellom denne reaksjonstype og hypnoseletthet (58, 63). Han kaller dette den triadiske hypotese.

Senere kommer Rosenzweig på ny inn på at glemsel av frustrerende opplevelser i betydelig grad kan forklares ut fra frustrasjonens egenart. Særlig når hindringen er slik at den direkte berører fpp.s stolthetsfølelse, vil de ha en tendens til å glemme de oppgaver de ikke har greid å løse<sup>1</sup> (61).

Mens Rosenzweig til å begynne med i første rekke synes å feste seg ved de individuelle avvik han fant på tvers av uttalte gjennomsnittlige gruppeforskjeller, og konsentrerte seg om å forklare disse avvik, blir han senere også interessert i å ta opp til drøfting de gjennomsnittlige forskjeller mellom gruppene.

Han henviser her i første rekke til de to gruppers forskjellige arbeidsbetingelser og innstillinger. I den gruppen som arbeidet under uformelle betingelser, fremkalte oppgavene stort sett liten spenning. Mens interessen i den ene gruppe hovedsakelig ble henledet på de foreliggende oppgaver, og det ikke å bli ferdig betydde svært lite utover en viss rest-spenning rettet mot de øyeblikkelige problemer, var interessen blant medlemmene i den andre gruppen ikke bare oppgaverettet, men også prestasjonsorientert, slik at den manglende fullføring uunngåelig ble oppfattet som et nederlag. I den førstnevnte gruppen var det særlig et *behovs-persistent* motiveringsgrunnlag, i den andre hovedsakelig av *ego-defensivt* motivasjonsmøster. At den siste gruppen fortrinnsvis husket fullførte oppgaver, kan sees som uttrykk for at et behov for *forsvar* her er blitt vakt, og at dette behov har tatt overhånd i forhold til en naturlig oppgave-spenning som skulle medføre hukommelse særlig for ufullførte oppgaver.

Ikke alle fpp. viste tegn på fortregning. Dette kan bero på at disse hadde relativt svake impunitiv tendenser, slik Rosenzweig først antok. Eller det kan bero på at disse ikke oppfattet aktivitetsblokkeringen som et personlig nederlag, at de hadde relativt svake ego-defensive tendenser.

Dette siste forklaringsprinsipp utelukker ikke gyldigheten av det første. De kan på sett og vis betraktes som komplementære. Det siste forklaringsprinsipp fører imidlertid til at en i tillegg til inndelingen mellom *reaksjonsretninger* også foretar en inndeling mellom to forskjellige *former* for reaksjonstendenser på frustrasjon.

En reaksjon kan enten ha sitt utgangspunkt i et enkelt segmentalt behov og ha som hovedfunksjon å tilfredsstille dette behov, eller den kan ha et motiveringsgrunnlag som omfatter et forsvar for personlighetens integrering, et forsvar overfor trusler mot denne integrering fremkalt av den frustrerende situasjon.

Den førstnevnte reaksjonsform, den «behovs-persistente», vil ifølge Rosenzweig fremkomme ved enhver frustrasjon medmindre den ikke blir

---

<sup>1</sup> Med bakgrunn i Zeigarniks berømte eksperimenter skulle en vente at fortrinnsvis uferdige og uløste oppgaver skulle virke ansporende og bli husket. Mangelen på fortregning i Zeigarniks materiale tillegger Rosenzweig i første rekke det forhold at oppgavene her ble gitt under uformelle betingelser. Var oppgavene gitt som prestasjonsprøver, ville resultatene sannsynligvis blitt annerledes.

trenget til side av en «ego-defensiv» reaksjonsform. I mange tilfelle vil disse to typer lett gå over i hverandre og ofte begge være til stede. Rent teoretisk vil imidlertid en avgrensning kunne være av verdi.<sup>2</sup>

I en artikkel i samleverket *Personality and Behavior Disorders* (1944) utvider Rosenzweig sitt klassifiseringsskjema til også å omfatte reaksjonenes form (64).

Han skiller mellom behovs-persistente og ego-defensive reaksjoner. Videre foretar han en tredeling av ego-defensive reaksjoner i forhold til retning: ekstrapunitiv, intro-punitiv og impunitiv ego-defensive reaksjoner, slik at han faktisk får å gjøre med fire atskilte reaksjonstyper. Han fremsetter den hypotese at de ego-defensive reaksjonsformer har en karakteristisk genetisk rangorden. Han anser de ekstrapunitive reaksjoner som de tidligste, dernest de intro-punitive som en konsekvens av en hemning av disse, og til slutt de impunitive som følge av en hemning av disse igjen.

Behovs-persistente reaksjoner kan være høyst forskjellige og behøver ikke være direkte. Både sublimering og substitutt-tilfredsstillelse gjennom symptomdannelse betrakter således Rosenzweig som behovs-persistente reaksjoner.

Så vel behovs-persistente som ego-defensive reaksjoner kan være mer eller mindre *adekvate*. En behovs-persistent reaksjon som blir gjentatt og gjentatt til tross for totalt uovervinnelige vansker for mål-oppnåelse, vil være en rigid, inadekvat reaksjon. Likeledes vil en intro-punitiv reaksjon være inadekvat, f. eks. hvis en person bebreider seg selv for nederlag som påviselig skyldes andres feilgrep.

Mennesker har forskjellig evne når det gjelder å mestre frustrerende situasjoner på en hensiktsmessig måte. Som en betegnelse på denne evne til å unngå inadekvate reaksjonsformer innføres begrepet *frustrasjonstoleranse*. Fra en psykoanalytisk synsvinkel, hevder Rosenzweig, kan frustrasjonstoleranse betraktes som et sidestykke til prinsippet om evne til utsatt tilfredsstillelse — en egenskap knyttet sammen med *jeg-styrke*. Jo mer jeg-svak et individ er, desto større behov vil det ha for å søke umiddelbar tilfredsstillelse og forsvare seg mot ubehag knyttet sammen med frustrasjon.

Jeg-svakhet vil med andre ord disponere for inadekvate reaksjoner på frustrasjon. En registrering av frustrasjonstoleranse skulle dermed legge grunnlaget til rette for en eksperimentell tilnærming til sentrale personlighetspsykologiske prosesser.

Spørsmålet om hvorledes en skal registrere frustrasjonstoleranse på en objektiv måte, reiser imidlertid store problemer.

---

<sup>2</sup> Praktisk talt samtidig med at Rosenzweig foretok sin inndeling mellom «ego-defensive» og «behovs-persistente» reaksjoner, foreslo Maslow (44, 45) en lignende inndeling — en oppdeling av frustrasjon i to komponenter — «personlighetstrusel» og «deprivasjon». Mens Rosenzweig i første rekke hadde sin oppmerksomhet rettet mot forskjellige reaksjonsformer på frustrasjon, tok Maslow utgangspunkt i en nærmere analyse av selve den frustrerende situasjon.

Som tidligere nevnt utarbeidet Rosenzweig i slutten av 30-årene en relativt hendig metode til registrering av ekstrapunitive reaksjonstendenser. Denne metode hadde imidlertid påfallende mangler. Rosenzweigs målsetting ble derfor å forbedre denne metode eller utarbeide en ny som ikke bare ga muligheter for registrering av ulike reaksjonsretninger, men også av forskjellige reaksjonsformer.

Rosenzweig valte å forsøke seg frem langs nye linjer. I 1944 tok han copyright på en billedassosiasjonsprøve, en prøve som senere i stadig større grad er blitt referert til som Rosenzweigs B-F-test. I 1945 ble en første beskrivelse og drøfting av testen publisert i *Journal of Personality* (65).

Under bearbeidingen av testen ble det klart at metoden også kunne anvendes overfor barn. I 1948 fulgte så en spesiell utgave av testen, beregnet for barn fra 4 til 13 år (70).

## Kapitel III

# Beskrivelse av testen

Rosenzweigs B-F-test ble introdusert som en billedassosiasjonsmetode, en metode som i sin form lå midt mellom ord-assosiasjon og tematisk appersepsjonsteknikk. På den ene side ligner testen TAT i bruken av bilder som stimulusmateriale for å fremme identifiseringsmekanismer. I motsetning til TAT er bildene stort sett like i sin komposisjon og blir nytt utelukkende til å fremkalle relativt enkle verbale svar. Mulighetene for en enkel og objektiv skåringsprosedyre er til stede. På dette punktet har testen uttalte likhetstrekk med en vanlig ordassosiasjonsprøve. I motsetning til en slik prøve tar testen imidlertid sikte på å favne mer molare sider ved personligheten.

Testen består av 24 ledd. Hvert ledd er et bilde, slike vi kjenner fra tegneserier, som viser to personer som er innviklet i en ofte forekommende mildt frustrerende situasjon. (Kfr. Appendix I.) Personen på venstre side av hvert bilde er blitt tillagt et utsagn som enten tjener til å beskrive frustrasjonen — eller representerer i seg selv et frustrerende element. Personen på høyre side av hvert bilde (subjektet) er alltid tegnet med en blank tekstrubrikk over seg.

For å fremme projeksjon er alle ansiktsuttrykk blitt fjernet, og figurene er blitt gitt et anonymt preg. Fpp. blir instruert om å ta for seg ett bilde om gangen, å skrive den aller første replikk som faller ham inn, i den åpne rubrikken på høyre side. Videre ble han i den opprinnelige utgave av testen anmodet om å unngå å være morsom, en anmodning som imidlertid senere er blitt fjernet fra instruksjonen, da den empirisk viste seg ikke å ha noen betydning (75).

De 24 bildene i testen fremstiller situasjoner av to forskjellige slag: *ego-blokkerende* og *superego-blokkerende* situasjoner. Det er ingen skarp grense mellom disse situasjonstyper, men de lar seg allikevel i prinsippet skille fra hverandre. *Ego-blokkerende* situasjoner er slike som fremstiller en personlig eller upersonlig hindring for måloppnåelse. *Superego-blokkerende* situasjoner (bilde nr. 2, 5, 7, 10, 16, 17, 19 og 21) derimot representerer ingen konkret hindring, men en utfordring, en beskyldning eller en anklage som blir rettet mot personen. De impliserer



i alminnelighet at en ego-blokkering allerede har funnet sted og at den frustrerende person tidligere har vært den frustrerte.

Testleddene blir skåret enkeltvis og hvert svar vurdert både med hensyn til reaksjonsretning og reaksjonsform. Etter at svarene på hvert bilde er blitt skåret, blir skårene summert.

Det klassifiseringssskjema som danner grunnlaget for skåringen, omfatter 6 *skåringskategorier*: 3 forskjellige reaksjonsretninger og 3 forskjellige reaksjonsformer, og 9 såkalte *skåringsfaktorer*, dvs. kombinasjoner av retninger og former. Skjema 3 gir en samlet oversikt over de forskjellige skåringsenheter, deres skåretegn og definisjon (kfr. 71, s. 169).

Mens Rosenzweig tidligere begrenset de tre reaksjonsretninger, ekstrapunitiv, intropunitiv og impunitiv, til den ego-defensive reaksjonsform, foretar han nå en lignende inndeling også av behovspersistente og *hindringsdominante* reaksjonsformer.

*Hindringsdominans* er betegnelse på reaksjoner hvor en konsentrasjon om hindringen eller blokkeringen er det mest fremherskende trekk. Hindringsdominans er en begrepsmessig nyskaping som såvidt en kan forstå selv tvang seg frem i forbindelse med utprøvingen og skåringen av testen. En rekke svar lot seg hverken klassifisere som ego-defensive eller behovspersistente, idet de hverken gikk i retning av en konsentrasjon om eget selv eller problemløsning, men mer eller mindre var låst fast i en konsentrasjon om hindringen som fremkalte den frustrerende situasjon.

Et par konkrete eksempler kan tjene til å illustrere skåringen.

Testsituasjon nr. 1 (kfr. s. 124) viser en bilist som ber en fotgjenger om unnskyldning for å ha sølt til klærne hans. Forsøkspersonen har som oppgave å svare på vegne av fotgjengeren.

Det er mange måter å ta en slik situasjon på. Noen er først og fremst innstilt på å bebreide og anklage andre, andre mer preget av selvbebreidelser og atter andre i første rekke innstilt på å tilgi og unnskyldde.

Nedenfor er angitt noen karakteristiske svar på dette bilde. Svarene er valt ut slik at alle de forskjellige reaksjonstyper er representert.

«Jeg beklager det også. Dressen er ny.» (E')

«Hvorfor ser De Dem ikke bedre for hvor De kjører.» (E)

«De burde påta Dem å få rensset dressen?» (e)

«Jeg ble ikke det spor tilsølt.» (I')

«Det er o.k. Det var min skyld som ikke holdt meg mer tilbake.» (I)

«Det er i orden. Jeg kan få rensset den.» (i)

«Dressen kan lett gjøres ren.» (M')

«Ikke noe å gjøre ved. Slikt hender.» (M)

«Den vil sikkert tørke på et øyeblikk.» (m)

Bokstavene i parentes illustrerer skåringen av de forskjellige utsagn.

Nedenfor er angitt noen karakteristiske svar på testsituasjon nr. 3. Denne situasjon viser to venninner som sitter på teater eller kino. Den ene bemerker til den andre som sitter bak en kvinne med en stor, bredbremmet hatt: «Du ser jo ingen ting.»

*Rosenzweigs reviderte klassifiseringsskjema, utarbeidet i forbindelse med B-F-testen*

| Reaksjons-retning     | Reaksjonsform  |  |  | Behovs-persistent (B-P)  |
|-----------------------|--|--|--|--|
|                       | Hindrings-dominant (H-D)   | Ego-defensiv (E-D)   |  |  |
| Ekstrapunitiv<br>(Ex) | (E') Nærværet av den frustrerende <i>hindring</i> blir sterkt fremhevet.   | (E) <i>Skyld</i> , fiendtlighet etc. blir rettet mot en person eller gjenstand i miljøet.<br>(E) Subj. <i>benekter</i> aggressivt sin <i>skyld</i> for den forseelse han blir anklaget.              |  | (e) Subj. venter at andre skal bringe en <i>løsning</i> på den frustrerende situasjon.   |
| Intropunitiv<br>(In)  | (I') Den frustrerende <i>hindring</i> blir tolket som ikke-frustrerende eller på en måte som fordelaktig; eller subj. understreker sin forlegenhet ved å medvirke til en annens frustrasjon. | (I) <i>Skyld</i> , klander etc. blir rettet mot subj.s eget selv.<br>(I) Subj. innrømmer sin <i>skyld</i> , men <i>benekter</i> enhver avgjørende feil ved å henvise til uunngåelige omstendigheter. |  | (i) Subj. tilbyr erstatning eller oppreisning til <i>løsning</i> av problemet.   |
| Impunitiv<br>(Im)     | (M') <i>Hindringen</i> i den frustrerende situasjon blir redusert nærmest til en fornektelse av dens eksistens.  | (M) <i>Skyld</i> for frustrasjonen blir unngått, situasjonen blir betraktet som uunngåelig, særlig blir det „frustrerende” individ frikjent.   |  | (m) Det blir gitt uttrykk for håp om at tid eller normalt forventede omstendigheter vil bringe en <i>løsning</i> på problemet; tålmodighet og konformitet er karakteristisk. |

Bokstavene i parentes er skårings-tegnene for de forskjellige svar-kategorier og -faktorer. Kombinasjon av tegn er av og til nødvendig for helt å kunne dekke enkelte svar. E og I er skåringsvarianter for E og I til bruk særlig i superego-blokkerende situasjoner. Disse varianter blir betegnet som benektelser («Denials»). I enkelte undersøkelser over testen er de blitt skilt ut og behandlet separat som en selvstendig skåringskategori på like fot med H-D, E-D og B-P.

Skårings-tegnene som er blitt angitt ovenfor for retnings- og formkategoriene, avviker noe fra dem som er blitt foreslått av Rosenzweig. Istedenfor H-D, E-D og B-P nytter Rosenzweig O-D, E-D og N-P, og istedenfor Ex, In og Im nytter han E, I og M. Da de tre siste tegnene er helt lik faktor-tegnene i E-D-kategorien, oppstår lett alvorlige misforståelser.

- «Nei. Jeg ser ikke det spor.» (E')
- «Nei. Enkelte tar aldri hensyn.» (E)
- «Be henne ta av hatten.» (e)
- «Jeg misliker å be henne ta den av.» (I')
- «Jeg skulle ha valt en annen plass.» (I)
- «La oss finne en annen plass.» (i)
- «Det er det samme for meg.» (M)
- «Hun er sikkert ikke oppmerksom på at hatten sjenerer.» (M)
- «Hun tar den sikkert av når filmen begynner.» (m)

Selvsagt forekommer det svar som er relativt vanskelig å skåre. Enkelte ganger kan det være ønskelig å nytte dobbeltskåring. De aller fleste svar byr imidlertid vanligvis ikke på særlige problemer. For å lette skåringen har Rosenzweig og medarbeidere dessuten utarbeidet egne skåringsutvalg. Det vil si at de har låst fast skåringen for en rekke flertydige svar som relativt hyppig forekommer, så skåringen på den måten ikke bare blir enklere, men også mer reliabel.

Lister over skåringseksempler og svarillustrasjoner foreligger både til testutgavene for voksne (68, 69) og barn (70).

En nærmere gransking av skåringsprinsippene (kfr. skjema 3) viser at både ego-defensive og behovs-persistente reaksjonsformer har fått et meningsinnhold noe forskjellig fra det Rosenzweig tidligere tilla disse betegnelser. Grunnen er nok så innlysende. Rosenzweigs tidligere klassifiseringsskjema var mer beregnet på å skape orden og oversikt over psykodynamiske prosesser og mekanismer innenfor rammen av et psykoanalytisk referansesystem enn å tjene som klassifiseringsgrunnlag for konkrete verbale reaksjoner.

Mens Rosenzweig tidligere så på forskjellen mellom ego-defensive og behovs-persistente reaksjoner som et uttrykk for hvorvidt et segmentalt behov eller personlighetens integrering ble søkt realisert, er tyngdepunktet nå lagt på hvorvidt en reaksjon er *problem-orientert* eller *skylds-orientert*.

Den nye betydning som blir tillagt ego-defensive og behovs-persistente reaksjoner, er i stor overensstemmelse med Newcombs senere inndeling mellom mål-orientert og trusel-orientert atferd (51), en inndeling som Newcomb på sin side i første rekke har hentet fra Maiers teoretiske skille mellom motivert og frustrert atferd (43).

I og med Rosenzweigs begrepsmessige endringer får også hans betegnelse *frustrasjonstoleranse* et nytt innhold. Mens han tidligere betraktet frustrasjonstoleranse som et individs evne til å unngå *inadekvate* behovs-persistente eller ego-defensive reaksjoner, definerer han nå frustrasjonstoleranse operasjonelt som *et individs tendens til å unngå ego-defensive reaksjoner*. Videre fremsetter han den hypotese at ego-defensive reaksjonstilbøyeligheter vil gi uttrykk for et individs jeg-styrke<sup>3</sup>, og behovs-persistente tilbøyeligheter, uttrykk for dets tilpassingsevne.

<sup>3</sup> Dette kommer klart til uttrykk i følgende referat fra testjournalen til en paranoid pasient: «Han anklager miljøet ved den minste provokasjon,

I senere arbeider synes Rosenzweig ikke lenger å opprettholde den ovenfornevnte definisjon av frustrasjonstoleranse. Istedenfor en ensidig konsentrasjon om ego-defens kategorien, legger han nå vekt på alle kategorier, og istedenfor en ensidig hyppighetsbetraktning er det i første rekke graden av positive og/eller negative avvik fra statistiske gjennomsnittsnormer som er avgjørende (72).

Ved siden av summeringen av enkeltskårer i forhold til kategorier og faktorer, med en fellesbetegnelse kalt *profiler*, gir enkeltskårene også grunnlag for beregning av en *gruppekonformitetstaksering*, og av *mønstre* og *tendenser*.<sup>4</sup> Også gruppekonformitetstakseringen og tendenser er blitt antatt å antyde noe om graden av frustrasjonstoleranse (72, s. 113).

*Gruppekonformitetstaksering* (*Group Conformity Rating* eller *GCR*) er en variabel som angir i hvilken grad et individs reaksjon viser overensstemmelse med andres reaksjoner. GCR er ment å gi grunnlag for vurdering av et individs sosiale tilpassing, for dets evne til konformitet med konvensjonell atferd i stress-situasjoner. Operasjonelt definert er GCR den prosent av et individs svar som viser overensstemmelse med visse skåringskriterier. Skåringskriteriene er empirisk bestemt. De er ett eller to svar ved hvert enkelt ledd som har vist seg å bli nyttet i flertallet av tilfelle og signifikant hyppigere enn andre svar i et normalutvalg.

Mer *mønstre* siktes til egenskaper ved skåreprofilene. Opprinnelig ble det lagt vekt på tre forskjellige typer av mønstre: 1) Den relative frekvens av de forskjellige retningsskårer uten hensyn til form. 2) Den relative frekvens av de forskjellige formskårer uten hensyn til retning. 3) En angivelse av de tre faktorer med størst hyppighet uten hensyn til hverken form eller retning.

Videre ble det lagt vekt på tre *tilleggs mønstre*. Disse refererte seg til hyppighetsberegninger i forbindelse med de understrekte faktorer: 1) Summen av E-svarene, 2) summen av I-svarene, og 3) totalsummen av disse to typer av svar. Hensikten var å få et mål på i hvilken grad de forskjellige situasjoner blir oppfattet som superego-blokkerende, og individet er tilbøyelig til å reagere med *benektelser*.

I manualen for testutgaven for barn opplyses det at de opprinnelige mønstre har vist seg lite egnet i klinisk arbeid. Derimot blir tilleggs mønstrene, de såkalte superego- eller *S-E-mønstre*, viet stor oppmerksomhet. Antallet av slike mønstre blir foreslått utvidet til 6. De tre nye som blir innført, er: 4) forholdet mellom E- og E-faktorene, 5) mellom I- og I-faktorene, og 6) summen av Im-kategorien og I-faktorene. De to første er ment å bringe oversikt over i hvilken grad den

---

undertiden ved å holde tilbake ethvert uttalt uttrykk for aggresjon, men ved denne sin reservasjon projiserer han det. Mesteparten av de ekstrapunitiv reaksjoner er av ego-defensiv type, noe som trolig står for en ekstrem jigsawhet.» (65, s. 21).

<sup>4</sup> Testens skåreskjema gir et visst inntrykk av hvorledes summeringen av enkeltskårer, beregningen av *mønstre* og *tendenser* finner sted. Kfr. Appendiks II, s. 132.

enkelte ego-defensive faktor er ladet med benektelses-svar, mens den siste har til hensikt å gi et samlet uttrykk for det totale antall svar som i en eller annen form innbefatter en frikjennelse for skyld.

Den tredje form for analyse av testresultater er beregning av *tendenser*. En tendens-beregning representerer en slags «sekvensanalyse». Det er en beregning av hvor konsistent individet er i sine reaksjoner. Dels er det ment å gi uttrykk for hvorledes individet er tilbøyelig til å reagere på egne reaksjoner, og dels å gi uttrykk for hvorledes individet har for vane å møte frustrasjon under akkumulerende grader av frustrasjon. Den grunnleggende antagelse er at «hvis et individ har bygd opp stabile og adekvate reaksjonsmønstre på frustrasjon, vil det møte hver situasjon i dens egen sammenheng og reagere på den på en egnet måte uten å bli forstyrret av rester fra sin tidligere atferd. Hvis på den annen side dets reaksjoner er ustabile, kan dette være uttrykk for konflikt, at dets atferd i senere situasjoner representerer en kompensasjon for, eller en reaksjon mot, dets tidligere reaksjoner ... I overensstemmelse med dette synspunkt vil tilstedeværelsen av tendenser, uten hensyn til spesiell kvalitet, implisere ustabilitet i reaksjonstiltøyeligheter og representere en indeks på lav frustrasjonstoleranse.» (78, s. 306).

Operasjonelt defineres en tendens som en signifikant forskjell i hyppighet av en skåringskategori eller -faktor i første og siste halvdel av testen. Til beregningen av forskjeller anbefaler Rosenzweig bruk av formelen  $\frac{a-b}{a+b}$ , hvor a er hyppigheten av faktoren eller kategorien i første halvpart, b i siste. Forat forskjellen skal sies å være signifikant<sup>5</sup>, må brøken ha en minimumsverdi på  $\pm .33$ , det vil si forholdet mellom de to deler må i det minste være som 1 til 2. En annen betingelse som må være til stede før en kan snakke om en tendens, er at den respektive kategori eller faktor omfatter tilsammen minst 4 skårer.

I alt er det mulig å arbeide med 15 tendensdimensjoner, 9 faktortendenser og 6 kategoritendenser, og hver av dem kan være enten positiv eller negativ.

Til slutt skal nevnes at testen tar relativt kort tid å administrere, for voksne ca. 15 min. og for barn ca. 20 min. Den kan nyttes både som gruppe- og individualprøve, som individualprøve både i skriftlig og muntlig form. For voksne er den også blitt nyttet som selvadministrert prøve.

Når en bruker testen på barn, er selvsagt alderen av vesentlig betydning. For barn under 8—9 år er en bundet til en muntlig administrasjonsform, og for barn under 9—10 år til en individuell administrering.

Testens enkle administrasjon, enkle og oversiktlige skåringsregler, den relativt korte tid som er nødvendig til både besvarelse og skåring — vanligvis ca. 1 time pr. testprotokoll — gjør at en meget forventningsfullt går testens data nærmere etter i sømmene.

---

<sup>5</sup> «Signifikans» må her betraktes som et «pseudostatistisk» uttrykk, idet samplingfordeling for «tendensfunksjonen»  $\frac{a-b}{a+b}$  ikke er antydnet.

## Kapitel IV

# Testens statistiske normer

Ved bruk av B-F-testen i psykodiagnostisk arbeid forutsettes vanligvis statistiske normer. Noen vurdering og tolkning av profiler, S-E-mønstre, tendenser og gruppekonformitetstaksringer kan vanskelig finne sted medmindre vi kan foreta sammenligninger med et normalmateriale. I forbindelse med GCR vil selve skåringskriteriene vi nytter oss av, være avhengig av hvilket normal-utvalg vi tar som utgangspunkt. Svaralternativer som er populære eller hyppige innenfor en sosial gruppe, vil ikke nødvendigvis være populære innenfor andre grupper. Også i forbindelse med beregning av tendenser er standardiseringsutvalgets sammensetning av avgjørende betydning.

Som opprinnelig normmateriale nyttet Rosenzweig seg av 100 normale voksne, 50 menn og 50 kvinner. Han foretok en inndeling etter kjønn, men utover dette var hans materiale udifferensiert (69).

Et normmateriale basert på bare 100 fpp. er naturligvis svært utilfredsstillende. I et senere arbeid har Rosenzweig publisert reviderte normer, basert på et betydelig større utvalg, i alt 460 normale voksne (77).

Dette utvalg består av henholdsvis 236 menn og 224 kvinner innenfor aldersgruppen 20—29 år. På grunnlag av foreløpige undersøkelser blir det hevdet at materialet også kan nyttes for aldersgruppen 30—39 år.

Normutvalget er ikke valt ut etter noen spesielle prinsipper med hensyn til representativitet. I utvalget av menn finnes faglærte og ufaglærte arbeidere, ferdige og uferdige collestudenter, medisinske studenter og psykologistudenter; i utvalget av kvinner finnes sykepleierskeelever, collestudenter, stenografer og medisinske studenter. Det gjennomsnittlige utdannelsesnivå for hele utvalget blir oppgitt til 13,5 år formell skolegang.

Utvalget kan vanskelig sies å være noe representativt utvalg for det anvendte aldersintervall; til det har det et for markert tyngdepunkt på et høyt utdannelsesnivå. Når det gjelder utvalgets sammensetning i forhold til slike faktorer som bosted, økonomiske forhold, sosiale og kulturelle forhold, gis det ingen nærmere opplysninger.

De statistiske normer for de enkelte faktorer, kategorier, S-E-mønstre og for GCR blir oppgitt i aritmetiske gjennomsnittsverdier og standardavvik.

Når det gjelder skåringskriterier for GCR, foreligger det ingen revisjon. Disse synes å bygge på det opprinnelige utvalg fpp. For en rekke ledds vedkommende, i alt 12, ble det her ikke funnet noe svar eller svarkombinasjon på to, som skilte seg ut med hensyn til popularitet.

Heller ikke i forbindelse med kriteriene for beregning av tendenser foreligger det noen revisjon. Rent prinsipielt vil en tendens som nevnt si at det foreligger en signifikant endring i reaksjonstype fra første til siste halvdel av testen. En forutsetning er imidlertid at testleddene i første og annen halvpart er balanserte, at de ikke mer eller mindre konsistent appellerer til forskjellige faktorer eller reaksjonsmåter. Denne forutsetning viste seg ved utprøvingen av testen ikke å være holdbar. For flertallet av fpp. i normalmaterialet viste det seg å foreligge en noe hyppigere bruk av hindringsdominante svar, og en noe sjeldnere bruk av ego-defensive, behovs-persistente og ekstrapunitive svar i siste halvdel av testen, sammenlignet med i første. For disse tendensdimensjoner ble minimumsverdien for en signifikant endring derfor forhøyet fra .33 til .50.

Noen statistiske normer når det gjelder hyppigheten av tendenser, er ikke blitt publisert. Det synes for øvrig her etter undersøkelser i forbindelse med testutgaven for barn — ikke å være antall tendenser som er av diagnostisk betydning, men om tendenser foreligger eller ikke (78, s. 313).

Ved siden av Rosenzweigs opprinnelige og reviderte normer foreligger et norm-materiale utarbeidet av Bernard (6). I alt omfatter Bernards utvalg 175 menn og kvinner. Majoriteten av utvalget består av college-studenter og «graduates» bosatt i New York City. Aldersmessig er utvalget sterkt begrenset, og også når det gjelder utdanningsnivå, er det en sterk forskyvning i forhold til lang skoletid. Tallmessig er Bernards utvalg en god del større enn Rosenzweigs opprinnelige utvalg, men betydelig mindre enn det som danner utgangspunkt for Rosenzweigs reviderte normer.

En vesentlig forskjell mellom Bernard og Rosenzweig er måten testen ble administrert på. Mens Rosenzweig nyttet seg av både individuell og gruppeadministrering, har Bernard konsekvent gjort bruk av testen som selvadministrert prøve.

Også når det gjelder skåringsprinsippene, foreligger det forskjeller. Som nevnt består Rosenzweigs E-D-kategori dels av svar som påpeker skyldsorientering i en eller annen form, dels av svar, de såkalte skåringsvarianter, hvor benektelse av egne feil er fremtredende (kfr. s. 18). Bernard skiller ut benektende svar som egen reaksjonsform. Istedenfor E-D-svar snakker Bernard konsekvent om henholdsvis «Denials» og B-O-svar. B-O-svar vil si svar som peker hen på en skyldsorientering («Blame Orientation»). I alt får Bernard derfor 4 forskjellige reaksjonsformer. Bakgrunnen for inndelingen synes i første rekke å være

semantiske overveielser og ønske om å arbeide med mest mulig operasjonelle betegnelser.

Også Rosenzweig har påpekt betydningen av å skille de to svar typer fra hverandre. Under S-E-mønstre foretar han særlige summeringer av benektelsessvar, men han synes for øvrig å se på de to svar typer som uttrykk for den samme ego-defensive disposisjon.

Ved siden av å beregne særskilte normer for kvinner og menn for de enkelte kategorier beregner Bernard også normer for tendenser og særskilte kjønnsnormer for de enkelte faktorer, for GCR og for skårekriteriene i forbindelse med GCR.

Når det gjelder normer for GCR, går han derfor et skritt videre enn Rosenzweig, idet han i tillegg til å skaffe seg mål på et individs konformitet med en normal-gruppe også tar opp spørsmålet om kjønnsforskjeller. Det vil si han gjør det mulig ikke bare å beregne et individs konformitet med en normalgruppe, men også med en normalgruppe av samme kjønn.

De skårekriteriene han beregner for GCR, skiller seg vesentlig fra Rosenzweigs. Dette er høyst forståelig så lenge han arbeider med atskilte utvalg av kvinner og menn. Men også når han beregner kriterier for sitt totalutvalg, er forskjellene påfallende. Bare i forbindelse med 10 av testleddene finner han ett eller to svar brukt signifikant hyppigere enn andre, og bare ved 7 av disse 10 leddene finner han full overensstemmelse mellom sine og Rosenzweigs kriterier.

Som nevnt fant Rosenzweig det ønskelig i forbindelse med enkelte tendensdimensjoner å kreve en minimumsverdi på .50 for at en tendens skulle kunne sies å være til stede. Også Bernard foretar en analyse med hensyn til svarfavorisering i første og siste halvdel av testen. Han konkluderer på grunnlag av sine data at det utelukkende er en økning i hindringsdominante og minskning i behovspersistente svar som krever en høyere minimumsverdi enn .33. Når det gjelder disse endringer mener han imidlertid det er ønskelig å sette minimumsverdiene høyere enn de Rosenzweig har foreslått, — helt opp på henholdsvis .82 og .64.

Både Bernards og Rosenzweigs overveielser i denne forbindelse bygger utelukkende på rent statistiske resonnementer. Tendenser er ment å skulle gi uttrykk for en bestemt type av intoleranse overfor frustrasjon. Viktigere enn å fastslå statistiske minimumsverdier basert på normalutvalg er det å finne frem til de verdier som eventuelt har en påvisbar diagnostisk verdi.

Det samme kan i høy grad sies om gruppekonformitetstakseringen. Den diagnostiske betydning av kjennskap til hvorledes et individ reagerer sammenlignet med et normalutvalg, vil være bestemt av bl. a. hva slags normalutvalg vi arbeider med, i hvilken betydning utvalget er normalt.

Når det gjelder vurderingen og tolkningen av et individs skåreprofil, vil det selvsagt være ønskelig å kjenne til hva som er vanlig reaksjonsmønster i den sosiale sammenheng individet befinner seg. Men heller ikke her vil en vinne stor innsikt utelukkende ved å følge statistiske



normer basert på et normalutvalg. En må bl. a. spørre seg i hvilken grad testresultatene er valide.

De statistiske normer Rosenzweig og Bernard har publisert, er basert på relativt store utvalg. Det er sannsynlig at normene vil gi oss holdepunkter ved bruk av testen blant studenter og andre med et høyt utdannelsesnivå, under nordamerikanske forhold, men utover dette er vi på usikker grunn. Vi står overfor spørsmål om testens miljøstabilitet. I den grad den skulle vise seg å være et miljølabilt instrument, vil Rosenzweigs og Bernards normer selvsagt ha en høyst begrenset anvendbarhet.

Vi vil senere komme nærmere inn på dette spørsmål i forbindelse med drøfting av testens differensieringsevne overfor sosiale grupper og i forbindelse med fremstillingen av våre egne data.

## Kapitel V

# Testens reliabilitet

Rosenzweigs B-F-test ble introdusert som en metode til å klargjøre personers måte å reagere på overfor frustrasjon. Testen inneholder 6 skåringskategorier, 9 skåringsfaktorer og en del indekser.

En kan spørre seg om kategoriene faktisk dekker empirisk påvisbare homogene variabler eller om de utelukkende er teoretiske konstruksjoner. Vi står her overfor spørsmålet om skåringskategoriernes *indre reliabilitet* eller *konsistens*. Når det gjelder indeksene, vil dette være en irrelevant problemstilling. Den form for reliabilitet som i første rekke vil komme på tale i forbindelse med indeksene, vil være den såkalte *retest-reliabilitet* eller *stabilitet* — hvorvidt skåreverdiene fluktuerer over tid. Også i forbindelse med kategoriene — enten vi ser på disse som variabler eller indekser — kan det være ønskelig å kjenne til denne reliabilitetsform.

Enda en form for reliabilitet vil det være viktig å få klarlagt: testens *skåringsreliabilitet*. Hvis ikke denne er tilfredsstillende, vil bl. a. våre konklusjoner med hensyn til de to førstnevnte reliabilitetsformer bygge på usikker grunn og være av tvilsom verdi.

## SKÅRINGSRELIABILITET

En undersøkelse av en tests skåringsreliabilitet kan finne sted på forskjellige måter. En måte er at en og samme forsøksleder skårer en serie med tester to eller flere ganger med et visst tidsintervall mellom hver gang. Et mål på skåringsreliabiliteten vil vi da få ved å registrere samsvaret mellom skåringene.

Denne fremgangsmåte er bl. a. blitt nyttet i Brun-Gulbrandsens skolehjemundersøkelse, det hittil eneste publiserte norske arbeid hvor B-F-testen er blitt anvendt (10). Brun-Gulbrandsen skåret sitt materiale 2 ganger med ca. ett års mellomrom. «Ved reskåringen var Fl. seg ikke bevisst å huske noen av de enkelte skåringene,» skriver han. «For hvert enkelt svar gikk vi frem på følgende måte: Vi gjorde oss opp en mening om skåren, sammenlignet så med den tidligere skåren, og dersom de to falt sammen, ble den beholdt. Dersom det ikke var tilfelle, ble de to

vurdert mot hverandre, og den ene ble som regel valt som endelig skåre. Resultatet var at i 12,2 % av tilfellene fant vi det nødvendig å forandre den opprinnelige skåren .... I betraktning av at det er en projektiv test, synes det.... å være en ganske bra overensstemmelse. Det er også grunn til å tro at tallet ville kunne minskes *vesentlig* med større øvelse.» (10, s. 178).

Det at en og samme person skårer det samme materiale 2 ganger, vil alltid innebære en viss sannsynlighet for at bevisste eller ubevisste hukommelsesprosesser kan gjøre seg gjeldende. Til det kommer at en person meget vel kan nytte seg av en og samme referanseramme for skåring uten at dette impliserer noen objektiv skåringsprosedyre. En objektiv skåringsmetode vil først og fremst si at det foreligger en viss intersubjektiv enighet, basert på bruken av eksplisitte skåringsregler.

Den vanligste form for undersøkelse av skåringsreliabiliteten er derfor å la det samme materiale bli skåret av to eller flere personer på uavhengig grunnlag og til slutt foreta en beregning av enigheten mellom dem. Det er blitt foretatt tre undersøkelser over B-F-testens skåringsreliabilitet etter disse retningslinjer.

Den mest omfattende er blitt foretatt av Clarke, Rosenzweig og Flemming (14). Utgangspunktet for undersøkelsen var å etterprøve og kontrollere nøyaktigheten av testens opprinnelige skåringsmanual. Denne ble utarbeidet i 1945 og publisert året etter (68). Etter å være tatt i bruk viste det seg imidlertid at den ikke på alle måter innfridde de forventninger en hadde satt seg, utvalget av skåringseksempler var høyst ufullstendig og i enkelte tilfelle også til en viss grad inkonsistent.

Clarke, Rosenzweig og Flemmings undersøkelse omfatter tre enkeltprosjekter. Først ble en gruppe på 82 journaler skåret uavhengig av 4 dommere. En tabulering av de forskjellige dommers skårer av samtlige bilder ga grunnlag for en sammenligning mellom 6 forskjellige par. For hvert par ble det så beregnet hvor mange bilder som var skåret likt. Graden av enighet viste seg å variere en god del fra par til par, fra 51 % til 83 % identiske skåringer, med et gjennomsnitt på 74 %.

En inngående leddanalyse viste at nesten halvparten av uenigheten skrev seg fra 7 av testleddene. Disse ble tatt opp til spesiell gransking med sikte på å finne frem til hva som hadde forårsaket uoverensstemmelsen. Det opprinnelige skåringsutvalg ble så utvidet, og i enkelte tilfelle ble også skåringskriteriene noe endret.

Da dette var gjort, ble neste prosjekt iverksatt. En gruppe på 136 testjournaler ble skåret uavhengig av to personer. Overensstemmelsen viste seg nå å være 87 %. På ny ble enkelte ledds skåringskriterier revidert og utvalget av skåringseksempler utvidet.

For endelig å undersøke effekten av disse endringer ble et tredje prosjekt satt i gang. 100 testjournaler ble skåret uavhengig av to personer. Etter 25, 25 og 50 journaler ble skåringen stoppet og resultatene sammenlignet og drøftet. De første 25 journaler viste en uenighet på 16 %, og de neste 25 og de siste 50 på 14 % hver. Overensstemmelsen for samtlige var 85 %.

Ennå viste det seg å være store forskjeller mellom leddene, men alle

viste enighet på minst 75 %. Clarke et al. konkluderer med at den skåringsreliabilitet de kom frem til ved tredje prosjekt, nemlig 85 %, sannsynligvis ville ha vært enda høyere hvis testen var blitt administrert på individuelt grunnlag. En enighet på 90 % ville i dette tilfelle ikke vært utenkelig. De avslutter med: «Det skulle være mulig for dem som nytter B-F-testen, å skåre svarene med en grad av reliabilitet som møter ethvert fornuftig krav til en kontrollert projektiv metode av denne type» (14, s. 370).

De konkrete resultater, endringene av skåringsutvalgene, ga grunnlag for en omarbeiding av den opprinnelige skåringsmanual. En revidert manual ble så publisert i 1947 (69).

Denne omfatter eksempler på svar hentet fra ca. 500 testjournaler. Det blir antatt at minst 50 % av alle svar gitt av normale voksne vil være å finne blant skåringseksempelene.

Clarke og medarbeideres resultat, den oppnådde enighet på 85 %, gir selvsagt ikke noe fullstendig bilde av manualens verdi. Den vesentligste innvending som kan gjøres mot undersøkelsen, er at det ikke ble nyttet nye skårere hver gang. En høy grad av enighet mellom to skårere er i seg selv ingen garanti for skåringsobjektivitet. Det kan lett innarbeides felles skåringsregler som ikke blir gjort eksplisitt.

Av særlig interesse er derfor reliabilitetsundersøkelsen til Franklin og Brozek (21). Den tar direkte utgangspunkt i den reviderte skåringsmanualen. Undersøkelsen bygger på et utvalg på 72 testjournaler som ble skåret uavhengig av to personer. I motsetning til Clarke et al. beregner Franklin og Brozek enigheten ved hjelp av korrelasjonskoeffisienter, idet de konsentrerer seg særskilt om de forskjellige skåringskategoriene. Høyest samsvar fant de ved skåringen av behovspersistens, lavest ved ego-defens. Følgende koeffisienter ble beregnet i forhold til de 7 kategoriene: ekstrapunitiv .86, intropunitiv .87, impunitiv .82, hindringsdominans .66, ego-defens .55 og behovspersistens .93. I forhold til GCR fant de for den ene halvdel av utvalget et samsvar på .81, for den annen halvdel et samsvar på .89.

Både i forbindelse med hindringsdominans og ego-defens er samsvaret begrenset, men betraktet under ett antyder allikevel Franklin og Brozeks resultater at den reviderte testmanualen er forholdsvis tilfredsstillende. Også Brun-Gulbrandsen gjorde bruk av den reviderte manualen. Tar vi også hans resultater i betraktning, gis det til en viss grad dekning for en konklusjon om at B-F-testen synes å kunne skåres med en reliabilitet som møter de krav en med rimelighet kan stille til en slik prøve.

## RETEST-RELIABILITET (STABILITET)

I alt er det blitt publisert tre undersøkelser over B-F-testens retest-reliabilitet. Om den ene av disse, foretatt av Rosenzweig, foreligger det svært sparsomme data, mens de to andre derimot er blitt meget utførlig behandlet.

I manualen til testutgaven for barn forteller Rosenzweig et sted at

han har funnet at utgaven for voksne har en høyst tilfredsstillende retestreliabilitet. De forskjellige kategorier, hevder han, har gitt korrelasjoner fra .60 til .80 (70, s. 145). Noen opplysninger om tidsintervall mellom første og annen gangs testing, om antall fpp. som ble anvendt, etc. blir ikke oppgitt. Dette er opplysninger som det er av den største betydning å kjenne til når en skal vurdere retestkorrelasjoner.

Den mest omfattende retestundersøkelse av B-F-testen er blitt foretatt av Schwartz (85). Undersøkelsen ble utført på et utvalg med 200 fpp. Tidsintervallet mellom de to testinger var en måned. Testen ble gitt under de samme betingelser ved begge anledninger, bortsett fra at administrasjonsformen var noe forskjellig. Ved siste testing, forteller Schwartz, ble det tatt hensyn til at fpp. hadde gjennomgått testen tidligere, og den ble nå introdusert med følgende bemerkning: «Jeg ønsker at dere besvarer denne testen uten hensyn til noen tidligere test-situasjon.»

TABELL 1

*Oversikt over Schwartz's retestkorrelasjoner*

| Reaksjonsretning: | <i>r</i> | Reaksjonsform:    | <i>r</i> |
|-------------------|----------|-------------------|----------|
| Ekstrapunitiv     | .83      | Hindringsdominans | .42      |
| Intropunitiv      | .65      | Egodefens         | .47      |
| Impunitiv         | .46      | Behovspersistens  | .67      |
|                   |          | Benektelse        | .50      |
| Gjennomsnitt      | .68      | Gjennomsnitt      | .53      |

Av interesse er at Schwartz finner at retningskategoriene oppnår høyere gjennomsnittlig korrelasjon enn formkategoriene. Schwartz tolker dette som uttrykk for at reaksjonsretningen stort sett er et mer konsistent kjennetegn ved individer enn hvilken form reaksjonene får. Likeledes legger han stor vekt på at behovspersistens oppnår høyest reliabilitet blant formkategoriene da denne kategorien etter hans mening har avgjort diagnostisk verdi.

«De middels til høye reliabilitetskoeffisienter på mange av skåringskategoriene får ytterligere betydning,» skriver Schwartz, «når skåringsresultatene blir sammenlignet og blir funnet å variere noe for forskjellige autoriteter, endog blant Rosenzweigs medarbeidere. Instrumentet er meget lovende, og fremtidig forskning kan med utbytte bli rettet mot dets validitet og mot ledd- og mønster-reliabilitetene . . . .» (85, s. 161).

I sin isolering av benektelsessvar er Schwartz i overensstemmelse med Bernard. Også Bernard har foretatt en retestundersøkelse av B-F-testen (6). Bernards utvalg består av 105 fpp. Videre opplyses det at tidsrommet mellom første og annen gangs testing varierte noe for de forskjellige fpp., men for alle lå tidsrommet mellom 3 og 9 måneder.

TABELL 2

*Oversikt over Bernards retestkorrelasjoner*

|    | H-D | B-O | B-P | Ben | M   |
|----|-----|-----|-----|-----|-----|
| Ex | .59 | .77 | .57 | .35 | .73 |
| In | .75 | .30 | .68 | .42 | .52 |
| Im | .49 | .60 | .55 |     | .69 |
| M  | .58 | .73 | .71 | .38 |     |

«Reliabilitetsverdiene for B-F-testen ligger godt an i sammenligning med skåringskategoriene til andre projektive prøver som f. eks. Rorschach», konkluderer Bernard, «men de er tilstrekkelig lave til å antyde at det bør utvises forsiktighet i forsøk på å forutsi et individs fremtidige atferd på grunnlag av B-F-testen alene. Spredningen i reliabilitetsverdiene antyder at den samme pålitelighet ikke kan bli tillagt de forskjellige skåringskategorier.» (6, s. 331).

En sammenligning mellom Bernards og Schwartz's data gir grunnlag for flere refleksjoner. En merker seg at noen påfallende forskjell mellom retnings- og formkategorier ikke er til stede i Bernards materiale. Likeledes at den gjennomsnittlige korrelasjon for retningskategoriene er den samme i de to undersøkelser, mens gjennomsittsverdien for formkategoriene er høyere i Bernards enn i Schwartz's materiale. Dette fører videre til at den gjennomsnittlige retestkorrelasjon for samtlige kategorier blir høyere hos Bernard, noe som er ganske påfallende i betraktning av at tidsintervallet mellom første og annen gangs testing her var betydelig lengre enn hos Schwartz.

En mulig forklaring er at Schwartz ved sin oppfordring til forsøkspersonene ved retestingen om å besvare testen uten hensyn til tidligere testing faktisk har oppmuntret forsøkspersonene til å svare forskjellig fra første gangs testing. Eller det kan gi uttrykk for forskjellen i administrasjonen av testen. Som tidligere nevnt nyttet Bernard frustrasjonstesten som selvadministrert prøve. Også en tredje mulighet kan tenkes, nemlig at det foreligger forskjeller i de anvendte utvalg. En retest-reliabilitet vil trolig alltid til en viss grad være avhengig av bl. a. utvalgets sosiale klasseforankring og posisjon.

Ved siden av å regne ut reliabilitetsverdiene for de enkelte kategorier og de enkelte faktorer foretok Bernard også en retest-beregning av GCR og summen av benektelse og B-O-kategoriene, dvs. Rosenzweigs E-D-kategori. Han fant her et samsvar på .69, mens han i forbindelse med GCR fant en reliabilitet på .45.

Det er ytterst vanskelig å vurdere retest-korrelasjoner ved personlighetsprøver. Viktigere enn stabiliteten over tid er selvsagt validiteten — og en høy validitet forutsetter ikke uten videre høy retestreliabilitet. I den grad vi ønsker å forutsi en meget skiftende egenskap, kan f. eks. en lav retest-reliabilitet være knyttet sammen med høy validitet.

Hvis vi antar at reaksjonsmønstre på frustrasjon er en relativt stabil egenskap ved individer, følger at vi skulle vente en relativt høy retest-reliabilitet. Nøyaktig hvor høy vil vi ikke kunne si noe om. Ved personlighetsprøver står vi nemlig nesten alltid overfor det forhold at den samme underliggende egenskap kan ha varierende fremtredelsesformer. Rent konkret vil dette si at en og samme kvantitative verdi på en faktor eller kategori ofte kan ha vidt forskjellig psykologisk betydning — og forskjellige verdier ofte den samme betydning — alt etter hvorledes individets skåremønster for øvrig ser ut. De forskjellige enkeltkategoriers og enkeltfaktorens retest-reliabilitet kan med andre ord være begrenset, uten at dette medfører en tilsvarende begrensning av prøvens anvendbarhet.

Også et par andre synspunkter blir i denne forbindelse ofte gjort gjeldende: «Hvis man skal undersøke det samme individ ved to forskjellige anledninger med enten den samme eller en alternativ form av en projektiv prøve, vil det alltid være nødvendig å overveie muligheten av at deltagelse ved den første anledning kan ha påvirket svarene ved den annen . . . . Og en undersøkelse av et individs ytelser ved forskjellige anledninger er ikke utelukkende en bestemmelse av et instruments reliabilitet eller stabilitet — slik tilfellet er for en psykometrisk test. En får nødvendigvis samtidig et mål på personlighetens konsistens.» (79, s. 219). Enhver personlighet, hevdes det videre, må ha en viss variabilitet i sine reaksjonsmåter hvis den skal opprettholde sin dynamiske karakter.

Sett på bakgrunn av disse synspunkter er det ytterst vanskelig å vurdere de reliabilitetskoeffisientene som er blitt funnet. Fra en mer psykometrisk synsvinkel kan en allikevel ikke unngå å påpeke den relativt lave retest-reliabilitet i forbindelse med GCR og «benektelser». Ønskeligheten av å skille ut benektelser som egen kategori synes, ifølge Bernards resultater, på ingen måte selvvinnlysene. Også GCR's reliabilitet er en god del lavere enn de tradisjonelle skåringskategoriene. At denne indeks alene er i stand til å gi noe særlig adekvat mål på «sosial tilpassing», er derfor høyst tvilsomt.

## INDRE RELIABILITET (KONSISTENS)

En fullstendig reliabel test vil — i psykometrisk betydning — si en test som ved umiddelbart etterfølgende målinger av det samme

individ eller fenomen vil gi nøyaktig de samme verdier. Det er en test som er fullstendig fri for målingsfeil.

Mer presist uttrykt er det selvsagt aldri en test som sådan, men testskårene som kan være mer eller mindre reliable; de kan i større eller mindre grad gi uttrykk for individets virkelige evner og disposisjoner.

Ut fra dette synspunkt ligger det nær å spørre etter samsvaret mellom oppnådd skåreverdi og sann skåreverdi. Et mål på dette kan en få ved å selvkorrelere testen. Blant de forskjellige metoder som er blitt foreslått i denne forbindelse, er det halveringsmetoden som uten sammenligning har vært mest brukt. Denne metode er likevel ikke bestandig egnet. Den gir ofte noe varierende resultater alt etter hvordan en deler testen, og i de senere år er det flere som har kritisert metoden (se f. eks. 33).

De metoder som i dag kanskje er mest anerkjent, bygger på variansresonnementer og legger vekt på testleddenes interkorrelasjoner. En av disse metoder er Richardson og Kuders rasjonelle ekvivalensmetode (16), en annen Hoyts variansanalytiske metode (33).

Hoyts metode forutsetter beregning av svarenes totalvarians og totalvariansens fordeling mellom ledd, personer og rest. Reliabilitetsformelen ser slik ut:

$$r_{tt} = \frac{(\text{variansen mellom personer}) - (\text{restvariansen})}{(\text{variansen mellom personer})}$$

Reliabilitetskoeffisienten gir, ifølge Hoyt, direkte uttrykk for prosenten av de registrerte skåreverdiers varians som skyldes de sanne skåreverdiers varians. Det lar seg lett avlede at jo større denne prosentetsats er, desto større vil samsvaret være mellom to sammenlignbare målinger av en og samme ting. Dette er også den vanlige psykometriske definisjon av reliabilitet.

Vi kan omskrive Hoyts formel:

$$r_{tt} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$$

$\sigma_e^2$  står her for restvariansen og  $\sigma_x^2$  for variansen mellom personer. Hvis vi løser ligningen med hensyn på restvariansen eller feilvariansen og beregner kvadratroten, får vi uttrykket for den såkalte standardfeil til målingen ( $\sigma_e$ ). Denne gir oss et absolutt mål på hvor stor innvirkning feilene har på nøyaktigheten av våre målinger. Vi blir med andre ord i stand til på grunnlag av en registrert skåreverdi å si noe om innenfor hvilke grenser det er sannsynlig at individets sanne skåreverdi ligger, om hvilken tillit vi kan ha til våre registrerte skåreverdier.

En av de mest omfattende undersøkelser som er blitt utført over Rosenzweigs B-F-test, er reliabilitetsundersøkelsene til Taylor (97). Av kontrollmessige grunner anvendte Taylor to forskjellige metoder, både Hoyts variansanalytiske og Jackson-Fergusons metode, en metode som er avledet av Richardson og Kuders rasjonelle ekvivalensmetode.



Som fpp. nyttet Taylor seg av to utvalg, hver på 130 voksne menn, som ble prøvd i forbindelse med personundersøkelser eller yrkesrettleiing. Taylor behandler utvalgene både hver for seg og som en samlet enhet.

TABELL 3

*Oversikt over Taylors reliabilitetskoeffisienter*

| Utvalg                  | N   | Ex  | In  | Im  | H-D | E-D | B-P |
|-------------------------|-----|-----|-----|-----|-----|-----|-----|
| Første utvalg . . . . . | 130 | .59 | .17 | .45 | .19 | .25 | .49 |
| Annet utvalg . . . . .  | 130 | .56 | .27 | .32 | .01 | .41 | .59 |
| Samlet utvalg . . . . . | 260 | .58 | .22 | .39 | .10 | .33 | .55 |

Reliabilitetskoeffisientene varierer, som en ser, helt fra .01 til .59; i det samlede utvalg fra .10 til .58. Gjennomsnittet i det samlede utvalg er .36.

«Med vanlige standarder,» skriver Taylor, «er mål med reliabilitet av denne størrelsesorden uskikket til vurdering av individer. Noen argumenter kan bli gitt for Ex (.58) og B-P (.55), men resten av kategoriene er under .40.» (97, s. 151).

En lav reliabilitet kan på den ene side bety at skåringsverdiene er unøyaktige, at de gir et upålitelig bilde av individets såkalte sanne skåreverdier, på den annen side at skåringskategoriene sannsynligvis ikke måler én, men flere variabler.

En kan spørre seg hvorvidt skåringskategoriene i det hele tatt er i stand til å skille på en signifikant måte mellom personer. Et direkte svar på dette kan en nemlig få ved å signifikansprøve forholdet mellom personenes varians og restvariansen. (35, s. 136).

Taylor tar også dette spørsmål opp til undersøkelse og finner at «Unntagen for H-D er F-forholdet signifikant på .05-nivået, Ex også på .01-nivået» (97, s. 151). Taylor går videre over til å drøfte hvilken grad av tillit som kan bli tillagt de registrerte skåreverdier. Hans resultater i denne forbindelse er som ventet ikke særlig løfterike. Han finner at en sann skåre på gjennomsnittet til Ex og In må ventes å korrespondere til registrerte skårer i høyeste og laveste kvartil henholdsvis i en tredjedel og i 44 % av tilfellene.

Taylor ser ikke bort fra den mulighet at testen overfor et mer forskjelligartet utvalg vil vise seg å være mer reliabel. Men det han i første rekke fester seg ved som forklaring på den lave reliabilitet, er en del særegenheter ved selve testen, først og fremst at mange av testleddene sterkt synes å favorisere svar i én skåringskategori fremfor andre. På ledd nr. 7 fant han at hele 92 % av utvalget ga Ex-svar, mens overhodet ingen f.eks. ga Im-svar på ledd nr. 19. En slik favori-

sering av bestemte svar på forskjellige ledd fører til at den optimale korrelasjon reduseres, prinsipielt på samme måte som reliabiliteten av en evnetest ville bli redusert hvis testleddene har liten forskjell i vanskelighetsgrad. Det at testen har et svært beskjedent antall ledd sett i forhold til at en ønsker å måle hele 6 forskjellige kategorier, vil videre aksentuere virkningen av den sterke svarfavorisering som er til stede.

En parallell til Taylors reliabilitetsundersøkelse av skåringskategoriene er Taylor og Taylors analyse av GCR's indre konsistens (96).

Utgangspunktet for undersøkelsen er Rosenzweigs påstand at GCR antyder et individs «kapasitet til konformitet med konvensjonell atferd under stress» (71, s. 172). Taylor og Taylor setter seg som mål å undersøke hvorvidt GCR virkelig kan sies å representere en kapasitet, i den forstand at «personer som tenderer å gi konvensjonelle svar i enkelte frustrerende situasjoner, også tenderer i samme grad å gi konvensjonelle svar i andre like frustrerende situasjoner» (96, s. 250).

Spørsmålet kan også formuleres slik: Hvorvidt foreligger det en positiv interkorrelasjon mellom GCR-ledd-skårer, i hvilken grad har GCR en indre konsistens?

For å besvare dette tar de i bruk Hoyts variansanalytiske metode.

De finner en reliabilitetskoeffisient for GCR på  $-.04$ . Og de konkluderer med at «de foreliggende data støtter ikke hypotesen om GCR's indre konsistens. Realiteten av et 'gruppekonformitetstrekk' blir ikke støttet, heller ikke kan GCR påvises å være et reliabelt mål på noe trekk . . . Det er ikke noe tegn for fremkomsten av en generell faktor gjennom positive interkorrelasjoner mellom leddene.» (96, s. 251).

Taylor og Taylor påpeker som en mulig forklaring at leddsvarene som blir skåret som konvensjonelle, viser sterk variasjon med hensyn til popularitet. Det er her av betydning å merke seg at de i sin beregning av GCR bruker Rosenzweigs skåringskriterier. Det at de ikke utarbeider disse på grunnlag av eget materiale, kan muligens ha virket til å øke popularitetsvariasjonen. Men at det også foreligger visse interne svakheter ved GCR, er ytterst sannsynlig, særlig i betraktning av Bernards lave retest-reliabilitet, enda særlige skåringskriterier var blitt utarbeidet. Både ved sammenligninger mellom forskjellige individer på samme tid og mellom samme individer på forskjellig tid synes GCR å ha en lite tilfredsstillende reliabilitet.

Taylors resultater viser at B-F-testens forskjellige kategorier har en meget begrenset indre konsistens. En kan imidlertid ikke dermed konkludere at testen er uten verdi. På samme måte som det er ytterst vanskelig å vurdere personlighetsprøvers anvendbarhet på grunnlag av deres retest-reliabilitet, er det heller ikke umiddelbart gitt hvilke konsekvenser en skal trekke av deres i alminnelighet ytterst lave indre konsistens (25, s. 100). Avgjørende vil alltid være prøvenes validitet.

Vi skal i de neste kapitler behandle B-F-testens validitet. I forbindelse med fremstillingen av egne undersøkelser vil vi igjen komme tilbake til spørsmålet om B-F-testens indre konsistens og forsøke å belyse spørsmålet ut fra en videre synsvinkel.

## Kapitel VI

### Testens validitet: 1. Generelle betraktninger

Som nevnt ble frustrasjonstesten opprinnelig konstruert som et eksperimentallpsykologisk hjelpemiddel. I de senere år er den imidlertid i stor utstrekning blitt brukt i psykodiagnostisk arbeid. Spørsmålet om testens validitet er blitt stadig mer fremtredende.

På samme måte som vi kan snakke om forskjellige reliabilitetsformer, kan vi også snakke om forskjellige typer av validitet. I U.S.A. er det i de senere år blitt vanlig å skille mellom fire slike typer, nemlig «*content*» validitet, «*predictive*» validitet, «*concurrent*» validitet og «*construct*» validitet (98). Å snakke om en tests validitet uten samtidig å presisere hva slags validitet en sikter til, vil ofte føre til uklarhet.

Med «*content*» validitet siktes til en tests representativitet, hvorvidt den gir grunnlag for å slutte noe til det univers av situasjoner som testen evt. utgjør et utvalg av. Når det gjelder B-F-testen, er vi ute av stand til å gi noe kvantitativt uttrykk for denne validitetsform.

Stort sett synes testens forskjellige situasjoner å gi et allsidig utvalg av «mildt frustrerende, ofte forekommende situasjoner». At vi gjennom testen faktisk stiller fpp. overfor slike situasjoner, er imidlertid en antagelse som må nærmere studeres. Før dette er gjort, vil spørsmålet om den innholdsmessige validitet vanskelig kunne drøftes.

Med «*predictive*» validitet siktes til en tests evne til å forutsi et individs fremtidige atferd, mens «*concurrent*» validitet reiser spørsmålet om testskårene gir et adekvat uttrykk for individets øyeblikkelige ytelser eller status. Begge former forutsetter at vi har uavhengige ytre validitetskriteriet, men tidsperspektivet er forskjellig. En test kan således godt ha «*concurrent*» validitet uten å ha noen «*predictive*» validitet.

Med «*construct*» validitet sikter vi til en tests evne til å gi oss opplysninger om indre disposisjoner, latente trekk eller kvaliteter. I dette tilfelle vil en validitetsundersøkelse i alminnelighet omfatte noe mer enn en rent empirisk analyse. Vi vil stå overfor det forhold at ytelser på testen peker hen på noe annet. Dette annet er av hypotetisk karakter. Våre tolkninger av skårer og skåremønstre vil dermed være bestemt av

vår teoretiske orientering. Den samme test vil kunne gi høyst forskjellige data alt etter hvilket teoretisk grunnsyn som danner bakgrunn for tolkningene.

Å bestemme en tests «construct» validitet vil i prinsippet si å vise at bestemte hypotetiske prinsipper eller konstruksjoner forklarer ytelser på testen. En validitetsundersøkelse vil derfor like mye være rettet mot teoretisk grunnsyn som mot selve testen.

En antagelse om f.eks. at lav frustrasjonstoleranse vil gi seg tilkjenne på B-F-testen som en sterk tilbøyelighet til å gi ego-defensive svar, vil aldri kunne prøves direkte. Vi vil imidlertid på indirekte måte kunne støtte eller svekke en slik hypotese — f.eks. ved å undersøke personer før og etter psykoterapi, ved å sammenholde våre slutninger med uttalelser fra klinisk trenede observatører, etc. Hvis det foreligger liten overensstemmelse med disse andre opplysninger, vil vi litt etter litt oppgi vår hypotese. Muligens vil vi gjennom de erfaringer vi har høstet, stille opp nye antagelser — f.eks. at antallet av tendenser gir uttrykk for graden av frustrasjonstoleranse, og at tilbøyeligheten til å gi egodefensive svar er tegn på f.eks. en autoritær skyldfølelse. På ny vil vi bli stilt overfor oppgaven å påvise at våre antagelser er holdbare. Og på ny vil det foreligge muligheter for at vi enda engang forandrer våre hypoteser og finner frem til andre og nye tegn som vi ikke tidligere har vært oppmerksom på.

På den annen side vil denne prosess ofte også føre til at vi modifiserer våre oppfatninger av den teoretiske konstruksjon vi ønsker å registrere. Selve valideringen vil derfor bestå i en gradvis høyere presisjon i vår evne til å slutte fra manifesterede tegn til latente trekk, samtidig som de latente trekk, våre hypotetiske konstruksjoner, blir mer og mer avklart.

En slik vekselvirkning mellom tegn og trekk, mellom empiri og teori, foregår kontinuerlig i klinisk arbeid hvor det foreligger muligheter for en eller annen slags kontroll av forutsigelser og diagnosestillinger. Bruk av f.eks. B-F-testen under slike betingelser fører nesten umerkelig til at det utformes en rekke mer eller mindre personlige tolkningsregler. Svært ofte vil disse være uverbaliserbare og funksjonere på et intuitivt, subliminalt plan. De forskjellige faktorens «ladninger» vil stå frem som tegn, og få seg tilknyttet meninger som på mange måter kanskje avviker sterkt fra de formelle betegnelser.

Når vi skal vurdere testens mer generelle anvendbarhet, må vi imidlertid i første rekke spørre etter hvilke eksplisitte regler som foreligger, og i hvilken grad disse har en påvisbar validitet. At testen brukt av en erfaren klinisk psykolog kan gi verdifullt diagnostisk materiale, sier mer om dens muligheter enn om dens aktuelle anvendbarhet.

Det er imidlertid all grunn til å legge stor vekt på tolkningsforslag fra den erfarne kliniker. På grunn av sin stilling vil han ofte ha særlige forutsetninger for å formulere sentrale tolkningshypoteser. Hans eget erfaringsmateriale vil vanligvis være for begrenset til å fastslå hypotesenes kvantitative gyldighet, men selve hypoteseformuleringen vil ofte kunne ha vidtrekkende betydning. De vil bl. a. kunne danne

utgangspunkt for kvantitative vitenskapelige analyser og dermed i vesentlig grad bidra til den nærmere klargjøring av testens anvendbarhet.

Det foreligger i dag, til tross for B-F-testens store utbredelse i de senere år, praktisk talt ikke noen samlet fremstilling av forslag til regler for tolkning av mønstre og enkeltfaktorer — basert på klinisk erfaring. Et hederlig unntak er en kort artikkel av Bernard (7).

Bernard advarer her bl.a. mot *bare* å tolke en høy skåreverdi på den ekstrapunitive kategori som uttrykk for utoverrettet aggressivitet. Ofte bør en høy ladning på den intro-punitiv-behovspersistente faktor tillegges samme betydning. I overensstemmelse med Rosenzweigs tidligere forslag mener han imidlertid at en påfallende lav B-P-skåre tyder på aktiv feiltilpassing.

«Innledende forsøk på å tolke *B-F-mønstre* på bakgrunn av hva kategorinavnene antyder, viste seg å være svært misvisende,» skriver Bernard. «Vi har ikke funnet en høy ego-defens-skåre noe hyppigere å være tilknyttet en «jeg-svakhet som krever spesielt forsvar» . . . . enn vi har funnet et uttalt avvik i de andre skåringskategoriene å ha en slik betydning . . . . den interesserte psykolog bør forsøke å kutte bort alle bibetydninger i retning av aggresjon i forbindelse med punitivbegrepene, unntagen i forbindelse med B-O-svar. Feiltolkninger av mønstre på grunn av kategorinavnene kan bli redusert ved å tenke på ekstrapunitivitet som *utover-orientering*, intro-punitivitet som *selv-orientering*, og impunitivitet muligens som *passiv-orientering*. Idet vi betrakter formkategoriene på samme måte, har vi *hindrings-orientering*, *skyld-orientering*, *mål-orientering* og benektelser.» (7, s. 341).

Bernards påpekning av farene forbundet med strengt å legge de forskjellige kategoriens ytre meningsinnhold eller navn til grunn for tolkning av skåremønstre berører sentrale spørsmål. En vesentlig grunn til at vi i det foreliggende arbeid stort sett har avholdt oss fra å oversette «tekniske» betegnelser, har sin bakgrunn nettopp i håpet om i minst mulig grad å fremkalle uønskede assosiasjoner.

Ved innføringen av tolkningsprinsipper blir som nevnt testens validering nøye knyttet sammen med valideringen av tolkningsprinsippene. Da disse aldri er gitt én gang for alle, men kan forandres etter som arbeidet skrider frem, åpner det seg her muligheter for systematiske rent empiriske tilnærmelser.

Rosenzweig er inne på en slik tilnærming i forbindelse med hva han kaller «en metode til validering ved etterfølgende kliniske forutsigelser» (76). Metoden går i korthet ut på at det først fremsettes hypoteser om hva prøven måler, dernest tas prøven i bruk, og til slutt sammenholdes resultatene med alle mulige andre data vedrørende det samme individ. Disse andre data vil imidlertid ikke ubearbeidet representere noe uavhengig kriterium. Det vil måtte analyseres, tolkes og flettes sammen med henblikk på å *gi grunnlag for forutsigelse av hvordan subjektet har svart på den uvaliderte prøve*.

Ved å konfrontere forutsigelsene med testresultatene vil hypotesene kunne bli støttet eller svekket. I det tilfelle hypotesene ikke blir støttet,

vil de måtte omformuleres og på ny tas opp til prøving gjennom nye kliniske forutsigelser. Gjennom serier med slike enkelt-eksperimenter basert på kliniske forutsigelser vil det være mulig, mener Rosenzweig, for den nye test gradvis å bestemme sin egen validitet.

Det vil også være mulig å avskrive en test praktisk validitet. Dette vil være tilfelle hvis forutsigelsene skulle ha en tendens til å bevege seg i en mer eller mindre ubestembar retning f. eks. i sirkel — hvis deres riktighet ikke skulle vise seg å øke over tid, men vise seg inkonsistente og stadig falle tilbake på tidligere svekkede hypoteser.

Metoden vil gjøre det mulig for den enkelte test å bestemme sin egen «mening» uavhengig av de antagelser som dannet bakgrunn for dens konstruksjon. «Siden erfaring antyder at mange personlighetstester, spesielt projektive prøver, har muligheter og begrensninger påfallende forskjellig fra dem deres skapere la vekt på, vil en metode som tillater en objektiv selv-bestemmelse, ha innlysende fordeler,» skriver Rosenzweig. «Med den foreliggende metode vil kvantitative så vel som kvalitative data bli anvendt på en objektiv måte i valideringsprosessen . . . Implisitte aspekter ved personligheten og endog den «totale personlighet» kan bli tatt hånd om . . . Den eneste begrensning ved fremgangsmåten som også garanterer dens objektivitet, er at forutsigelsene må være formulerbare.» (76, s. 509).

Den metode Rosenzweig foreslår, kan med mindre endringer også anvendes utenfor den kliniske situasjon. Det vesentlige vil alltid være at hypotesene er prøvbare. Om en tar utgangspunkt i en enkelt person eller i grupper av personer, om en går fra testresultat til observerbar atferd eller fra atferd til testresultat, vil være av sekundær betydning.

Rosenzweigs forslag at en test selv skal bestemme sin egen validitet, forutsetter at en til enhver tid har klart formulerte hypoteser basert på hva andre har funnet at testen måler. En slik form for validering vil selvsagt måtte strekke seg over et langt tidsrom.

Vi kan naturligvis i dag ikke gi noe endelig svar på hva B-F-testen måler. Vi må begrense problemstillingen til hva vi i øyeblikket tror den måler. Dette kan vi besvare ved systematisk å gå igjennom resultatene fra samtlige validitetsundersøkelser som er blitt foretatt over testen. Ved en kritisk gjennomgåelse av andres undersøkelser vil vi ikke bare få en oversikt over hvilke hypoteser som er blitt støttet og hvilke som er blitt svekket, vi vil muligens også komme på spor etter nye hypoteser og få øye på nye muligheter for å undersøke testens validitet.

Før vi gir oss i kast med denne oppgave, er det et forhold vi finner det nødvendig å trekke frem. Nesten alle undersøkelser som hittil har vært foretatt over B-F-testen, har vært rettet mot dens «concurrent» validitet. Svært ofte har oppmerksomheten utelukkende vært rettet mot B-F-testens forskjellige kategorier, og disse har i stor utstrekning blitt oppfattet som ekvivalente med dimensjoner i manifest atferd. Først i de senere år har spørsmålet om B-F-testen måler latente trekk og disposisjoner, begynt å innta en fremtredende plass. Dermed har det også funnet sted en endring i synet på kategoribetegnelser og

adekvat validitetsform. Istedenfor forsøk på å finne ytre direkte atferdskriterier er mer indirekte tilnærmelser blitt foretrukket. Ennå synes det særlig å være de enkelte kategorier valideringsarbeidet blir konsentrert om, men også på dette område finner det sted en utvikling — en forskyvning av interessen fra enkeltkategorier til mønsteranalyser. Å utarbeide statistiske metoder som tar hensyn til mønstre, som gjør det mulig for oss å ta hensyn til det forhold at en og samme skåreverdi kan bety forskjellige ting, og forskjellige skåreverdier det samme, alt etter hvilken sammenheng de fremkommer i, reiser ytterst kompliserte problemer. På tross av dette har det faktisk i de aller siste år blitt utarbeidet metoder som kan være oss til hjelp nettopp til løsning av hva vi kan kalle mønster- eller profil-analyser (32). Et mindre, nylig publisert arbeid (99) over tre TAT-variabler med utgangspunkt i Rao's «multiple diskriminant teknikk» gir antydninger om utviklingslinjer på dette område. Rosenzweigs B-F-test er ennå ikke blitt undersøkt på denne måte, men testen skulle ligge meget godt til rette for eksperimentelle, statistiske analyser etter disse nye retningslinjer. Den har et relativt begrenset antall variabler, den er relativt hurtig å administrere, hurtig å skåre, etc.

Et spørsmål som fort reiser seg når en skal gå i gang med mønsteranalyser, er hvilket forhold de forskjellige skåringskategorier står i til hverandre. I hvilken grad er de uavhengige av hverandre? B-F-testen er slik konstruert at antall svar innenfor retnings- og formkategoriene er fiksert. En økning i én kategori vil samtidig redusere antallet i en av de to andre retningskategoriene, og hvis vi arbeider med benektelse som egen form — i en av de tre andre formkategoriene. Korrelasjoner blant retnings- og formkategoriene vil derfor i høy grad være uekte. Også korrelasjoner mellom retnings- og formkategorier vil ha begrensninger i den grad vi ser på «benektelse» som egen kategori, men begrensningene vil være mindre enn i første tilfelle.

Vi har i dag empirisk materiale som belyser forholdet mellom skåringskategoriene. I en undersøkelse av Schwartz (85) beregnes bl.a. skåringskategoriernes interkorrelasjoner. Schwartz skiller ut benektelsesvar som egen kategori, og innfører som en 8. kategori «antall svar». Han finner en positiv korrelasjon mellom ego-defensive og ekstrapunitive tendenser, og en negativ korrelasjon mellom ego-defensive og intropunitive tendenser. Antall svar finner han er signifikant positivt korrelert med intropunitive tendenser og signifikant negativt korrelert med behovspersistens. Likeledes finner han høy korrelasjon mellom intropunitive og behovspersistente tendenser.

På grunnlag av interkorrelasjonene mener Schwartz det er mulig forsøksvis å tolke hva testkategoriene gir uttrykk for. Interkorrelasjonen mellom antall svar, intropunitive og behovspersistente tendenser antyder, ifølge Schwartz, «tilstedeværelsen av en innstilling av samvittighetsfull hyperaktivitet, kanskje av selvavvergende karakter» (85, s. 159). Benektelse-kategorien mener han hypotetisk kan tolkes som uttrykk for «usikkerhet og frykt for gjengjeldelse fra omgivelsenes side», og lite antall svar som uttrykk for «sterk selvopptatthet, mangel

på hensyntagen til andre, og kanskje tendenser til ekspulsiv og impulsiv atferd».

Schwartz's tolkninger representerer empirisk prøvbare hypoteser. I seg selv sier de intet om testens anvendbarhet. Det er resultatene av empiristiske prøvinger som her vil være avgjørende.

Vi skal i de neste kapitler gå nærmere inn på et stort antall undersøkelser som er blitt foretatt med sikte på å klargjøre frustrasjons-testens validitet. For å gjøre fremstillingen mest mulig oversiktlig har vi valt å ordne dette materiale i tre kapitler. Vi vil først ta opp til drøfting testens differensieringsevne, dernest gi en oversikt over undersøkelser hvor testens samsvar med ytre atferdskriterier og andre testresultater er blitt studert, for til slutt å gå nærmere inn på spørsmålet om på hvilket *atferdsnivå* testen har sin forankring.



## Kapitel VII

# Testens validitet: 2. Differensieringsevne

Et fellestrekk ved de undersøkelser som skal behandles i dette kapitel, er at de har tatt utgangspunkt i en sammenligning av forskjellige grupperes testresultater. Er testen i stand til å differensiere mellom f.eks. forskjellige sosiale grupper? I den grad den er sensibel overfor gruppeforskjeller, blir det neste problem: Gir forskjellene i testresultatene noen mening, er de i overensstemmelse med hva vi skulle vente ut fra personlighetspsykologisk teori eller «common sense». Undersøkelsene belyser alle testens validitet. Men de vil også kunne gi oss opplysninger som kan være av betydelig verdi ved senere tolkning av testresultater og skåreprofiler. De vil kunne gi oss holdepunkter når det gjelder å formulere nye hypoteser om hva testen måler, og dermed tjene som utgangspunkt for senere empiriske undersøkelser.

### TESTENS DIFFERENSIERINGSEVNE OVERFOR FORSKJELLIGE SOSIALE GRUPPER

I hvilken grad reagerer *kriminelle* forskjellig fra andre? Dette spørsmål har lenge opptatt både psykologer og sosiologer. Det er blitt anvendt de mest forskjelligartede metoder, men ennå er spørsmålet langt fra løst. Svært ofte blir det antatt at kriminelle er mer aggressive enn ikke-kriminelle, at de er mer ukontrollerte og mindre preget av moralske hemninger.

For å kaste ytterligere lys over forholdet mellom kriminelle og ikke-kriminelle foretok Fry (23) en spesiell undersøkelse med B-F-testen. Han benytter seg av tallmessig relativt store utvalg. Utvalget av kriminelle består av mannlige og kvinnelige fengselsfanger, utvalget av ikke-kriminelle, av studenter av begge kjønn.

Første trinn i undersøkelsen er en sammenligning av de to utvalg med Rosenzweigs opprinnelige norm-materiale. Som en kunne vente, finner han størst samsvar mellom dette og studentutvalget. Fry begrenser for øvrig sin statistiske analyse til bare å omfatte sitt eget empiriske materiale. Ved siden av å sammenligne sine to utvalg foretar han

også en oppdeling av sitt materiale i forskjellige subutvalg — etter kjønn, rase og veteran status.

Størst forskjell mellom gjennomsnittsverdier på skåringskategoriene viser det seg å være mellom kvinner og menn, og mellom fanger og studenter. Men også mellom veteraner og ikke-veteraner og mellom negrer og hvite finner han enkelte signifikante forskjeller.

Han finner at studentene reagerer mer ekstrapunitivt og mindre intropunitivt enn fangene. Dette er kanskje ikke akkurat hva en skulle vente.

En mulig forklaring får en imidlertid ved å ta utgangspunkt i Rosenzweigs hypotese at høy intropunitivitet vil være knyttet sammen med «aggressive behov rettet innover» og «behov for straff». Fangenes aktuelle sosiale situasjon gir nettopp uttrykk for at de er blitt straffet. Hvorvidt de faktisk har hatt dypereleggende behov for å komme i en slik situasjon, kan ikke B-F-testen gi noe sikkert svar på. Resultatet, den høye intropunitivitet, kan imidlertid tolkes som uttrykk for et slikt forhold.

En annen forklaringsmåte er at testsituasjonen for de to utvalg var forskjellig. Fangene ble prøvd i fengsel, i et miljø hvor ekstrapunitive reaksjoner er tilknyttet sterke negative sanksjoner. Ofte kan det tillike være slik at reaksjonstendenser i ikke-ekstrapunitiv retning kan bringe en fange lettelse og særlige fordeler. Det er ikke utenkelig at mange av fangene bevisst eller ubevisst har nyttet testen til å demonstrere sin konformitet med fengselmiljøets krav.

Hvilken av de to forklaringer som er den rette i den foreliggende sammenheng, er det umulig å si. En ganske nylig publisert undersøkelse av Lindzey og Goldwyn — hvor det gjøres bruk av barneutgaven av testen — antyder at testsituasjonen er av særlig betydning (42). Lindzey og Goldwyn foretar en sammenligning mellom en gruppe på 40 skolehjemselever og 50 ikke-kriminelle barn — alle i 12—13 års alderen og alle med den samme sosialøkonomiske bakgrunn. Også de finner at kriminelle skårer signifikant lavere enn ikke-kriminelle når det gjelder ekstrapunitive reaksjoner. De går imidlertid et skritt videre, idet de nettopp for å undersøke testsituasjonens betydning foretar en nærmere sammenligning mellom de av skolehjemselevene som viser best og de som viser dårligst tilpassing til skolehjemmets krav om ikke-aggressiv atferd. De finner en signifikant forskjell mellom de to ekstremgruppers tilbøyelighet til å gi ekstrapunitive svar på testen. Elevene som viser best tilpassing, får gjennomgående en mye lavere ekstrapunitiv skåre enn de som viser dårligst tilpassing. Lindzey og Goldwin tolker resultatet som uttrykk for en generalisering av hemninger overfor ekstrapunitiv atferd som følge av skolehjemsinstitusjonens belønningssystem (42, s. 527). At det også i et fengselmiljø kan finne sted en slik generalisering av hemninger, synes høyst sannsynlig.

Når det gjelder forholdet mellom studenters og fangers reaksjonsformer, finner Fry størst forskjell i forbindelse med intropunitive tendenser. Mannlige studenter viser signifikant større tendens til behovspersistente reaksjoner enn mannlige fanger, og kvinnelige fanger signi-

fikant større tendens i denne retning enn kvinnelige studenter. Det utslagsgivende forhold synes å være kvinnelige fangers sterke behovspersistente tilbøyeligheter, og ikke det at kvinnelige studenter er uttalt svake. Forskjellene mellom mannlige og kvinnelige studenters reaksjonsformer synes over det hele stort sett å være små.

Når det gjelder kjønnsforskjeller, merker en seg spesielt at både blant fangene og studentene viser menn signifikant mer ekstrapunitiv tendenser.<sup>6</sup>

Ved siden av å utarbeide normer for de forskjellige skåringskategorier foretar Fry også en standardisering av GCR. Han finner ingen signifikante kjønnsforskjeller på denne indeks, men derimot en påfallende signifikant forskjell mellom studenter og fanger. Blant både menn og kvinner får studentene gjennomgående en betydelig høyere GCR enn fangene.

Hvis vi på ny foretar en sammenligning med Lindzey og Goldwyns resultater med barneformen av testen, blir vi slått av en bemerkelsesverdig forskjell. Lindzey og Goldwyn finner absolutt ingen forskjell i GCR mellom kriminelle og ikke-kriminelle barn. Derimot finner de en signifikant forskjell mellom de to utvalgs tendens-frekvenser (42, s. 525). Deres resultater er her i overensstemmelse med Rosenzweig og Mirmows data (78). I begge tilfelle blir det påvist et positivt samsvar mellom inadekvat sosial tilpassing og antall tendenser, men i begge tilfelle blir det gjort bruk av barneutgaven av B-F-testen.

Når vi skal vurdere Frys resultater, står vi til en viss grad på usikker grunn. Fry arbeider med utvalg av kriminelle og ikke-kriminelle, men resultatene kan ikke entydig tolkes som uttrykk for denne forskjell. Ved siden av å være forskjellig med hensyn til sosial status er utvalgene sannsynligvis også høyst forskjellig med hensyn til gjennomsnittlig utdannelsesnivå, sosial-økonomisk bakgrunnsmiljø, etc.

I hvilken grad er testen sensibel overfor slike forskjeller? Hvis dette i høy grad skulle være tilfelle, vil det innvirke ikke bare på tolkningen av Frys resultater, men også mer generelt, på anvendbarheten av Rosenzweigs norm-materiale.

En undersøkelse av McCary (46, 47) belyser dette nærmere. McCarys problemstilling var å undersøke hvorvidt reaksjonstendenser på frustrasjon er avhengig av etisk-kulturelle forhold. Hans hovedinndeling i denne forbindelse er kjønn, rase og bosted. Når det gjelder bosted, arbeider han med to grupper, én gruppe bosatt i nordlige og én i sørlige områder av U.S.A. Undersøkelsen omfatter i alt 631 personer i aldergruppen 14 til 22 år.

McCarys resultater viser at det foreligger store forskjeller mellom grupper fra nord- og sørstatene. Gruppen fra nordstatene er betydelig mer åpent aggressiv og mindre passiv enn gruppen fra sørstatene.

---

<sup>6</sup> Dette er i uoverensstemmelse med kjønnsforskjellene i Rosenzweigs opprinnelige normmateriale, men utvilsomt i overensstemmelse med vanlig populær oppfatning.

Kjønnsforskjeller viser seg å være minimale. Derimot finner han negrer generelt å være mindre passive enn hvite. Men til tross for denne generelle tendens viser det seg at blant menn i sørstatene er hvite mer aggressive og mindre passive enn negrer. McCary tolker dette som en bekreftelse på Dollards observasjoner — at sørstatnegrer er dominert av hvite, og at deres aggresjon som følge av dette er blitt omformet til en underdanig holdning.

Ved sin påvisning av store forskjeller mellom kulturelle og rase-messige grupper påpeker McCary indirekte begrensninger ved de statistiske testnormer som Rosenzweig har lagt frem. Ved siden av å ha en sterk geografisk-kulturell begrensning er det særlig det høye gjennomsnittlige utdannelsesnivå som fremkaller tvil om normalmaterialets anvendbarhet.

Vanligvis er det rimelig å regne med et visst samsvar mellom utdannelsesnivå og sosial-økonomisk status. En analyse av forholdet mellom sosial-økonomisk status og reaksjonstendenser på B-F-testen kan derfor muligens belyse spørsmålet.

«Ingen forsøk ble gjort,» opplyser McCary, «på å likestille forsøkspersonene på basis av sammenlignbarhet i sosial-økonomisk status, da det i forutgående undersøkelser med B-F-testen viste seg å være lite samsvar mellom individers sosialøkonomiske status og deres reaksjoner på frustrasjon» (46, s. 321). Disse forutgående undersøkelser utgjorde en beregning av korrelasjonen mellom skåringskategoriene og sosial-økonomisk status — basert på *Sims Score Card* for et utvalg på 94 college-elever.

En tidligere undersøkelse av Falls og Blake på et utvalg av 25 studenter fant heller ikke noen samsvar på dette punkt (18).

Et fellestrekk ved McCarys og Falls og Blakes undersøkelser er den høye grad av utdannelsesmessige homogenitet som preger deres utvalg. Dette medfører en sterk begrensning i de konklusjoner som kan trekkes av deres resultater.

En mer adekvat tilnærming for å undersøke samsvaret mellom testresultater og utdannelsesnivå vil sannsynligvis være å ta utgangspunkt i resultatene på en intelligens-test, da det i alminnelighet må antas å være et betydelig samsvar mellom intelligens og utdannelsesnivå.

Karlin og Schwartz har nylig publisert en undersøkelse hvor de bl.a. tar opp forholdet mellom intelligens og svartendenser på B-F-testen (36). Som intelligens-test benytter de seg av *Otis Test of Mental Ability*, og som forsøkspersoner 45 mannlige studenter. Også i dette tilfelle er utvalget med forsøkspersoner altså høyst ensartet. Deres resultater er ikke desto mindre av stor interesse. De påviser at det ikke foreligger noe samsvar mellom retningskategoriene og generell intelligens, men derimot at det er et signifikant positivt samsvar mellom intelligens og tilbøyeligheter i retning av hindringsdominante og behovspersistente svar.

Karlins og Schwartz's data må sies å gi sterke antydninger om at Rosenzweigs testnormer har begrenset gyldighet, og at de bør nyttes med stor forsiktighet.

Men også i forhold til Frys resultater gir Karlin og Schwartz's data visse holdepunkter. At det blant ikke-kriminelle menn er en høyere behovspersistens enn blant kriminelle, synes altså å være usikkert; derimot må de kvinnelige fangers høyere behovspersistens tillegges særlig vekt. Når en samtidig tar hensyn til de kvinnelige fangers høyere intro-punitivitet, ligger det nær å anta at det her særlig er den intro-punitiv-behovspersistente faktor som er utslagsgivende. Det er av interesse å merke seg at nettopp denne faktor av Bernard er blitt antatt å gi uttrykk for tendenser til utoverrettet aggressivitet (kfr. s. 37). Vi kan ennå ikke trekke noen sikre slutninger om dette, men det åpner seg her et spesielt problem hvor ytterligere prøving vil være ønskelig.

## TESTENS DIFFERENSIERINGSEVNE OVERFOR GRUPPER MED FORSKJELLIG KLINISK DIAGNOSE

Det er blitt foretatt to undersøkelser med særlig henblikk på å fastslå B-F-testens kliniske diskrimineringssevne.<sup>7</sup> Disse undersøkelsene er høyst relevante når det gjelder å belyse sider ved testens kliniske anvendbarhet.

I en undersøkelse av Seward og Morrison forsøker en å analysere personlighetsstrukturen til pasienter med en bestemt form for tykk-tarmsbetennelse — spastic colitis (87). Denne sykdom antas ofte å ha et psykosomatisk grunnlag. Ved siden av B-F-testen anvender Seward og Morrison også Rorschach og TAT.

Deres utvalg av fpp. består av 21 menn med en gjennomsnittlig alder på ca. 30 år.

Det mest bemerkelsesverdige trekk ved colitis-pasientene synes å være deres ualminnelig svake ego-defensive og uvanlig sterke behovspersistente tilbøyeligheter. Hele 67 % av colitis-utvalget oppnår en E-D-skåre i laveste kvartil på Rosenzweigs opprinnelige normmateriale, og hele 81 % en B-P-skåre i høyeste kvartil. Med hensyn til reaksjonsretning er det særlig deres lave ekstrapunitivitet og intro-punitivitet — til fordel for høy impunitivitet som er påfallende. Fremtredende trekk ved colitis-pasientene synes med andre ord å være en tendens til målorientering — eller kanskje riktigere ønsketenkning — hvor problemløsning blir utsatt til fremtiden og antatt å finne sted mer eller mindre av seg selv. Videre en sterk tendens til å unngå personlig innvikling og en manglende evne til å objektivisere mellom-menneskelige forhold. «I henhold til Rosenzweigs tolkninger,» skriver Seward og Morrison, «antyder resultatene at colitis-pasientene mangler tilstrekke-

<sup>7</sup> Ganske nylig er det blitt publisert en tredje undersøkelse med samme formål (9). Ved sammenligning mellom grupper av normale, kroniske alkoholikere og paranoide schizofrene påviser Brown og Lacey at den ekstra-punitiv og impunitiv kategori er i stand til å differensiere mellom utvalgene med en pålitelighet på under 1%-nivået og hindringsdominans på under 5%-nivået. Det er imidlertid stor overlappning mellom gruppene, og Brown og Lacey konkluderer med at forskjellene ikke er tilstrekkelig store til å berettiggte bruk av testen til individuell diagnose av patologiske tilstander (9, s. 75).

lig jeg-styrke til å mestre frustrasjon på et modent realitetsplan . . . .» (87, s. 17).

Også i Rorschach og TAT kommer det klart frem tendenser til å unngå emosjonelle opplevelser, i TAT dessuten sterke indikasjoner på at sosialt samvær er tilknyttet ubehag og at utlevelse og opplevelse av sinne og fiendtlighet er sterkt konfliktladet. I Rorschach-protokollene finner de tegn på en høy grad av funksjonell rigiditet, overfladisk ambitiøs strebing og mangel på spenningstoleranse. I det store og hele må overensstemmelsen mellom resultatene på de tre personlighetsprøver — etter Seward og Morrisons mening — sies å være påfallende god.

Mangel på frustrasjonstoleranse kommer i B-F-testen frem ved at en stor del (62 %) av utvalget viser tendenser. Hyppigst er tendenser som ga uttrykk for en endring i svartilbøyeligheten bort fra behovspersistens og intropunitivitet.

Seward og Morrison foretar også en sammenligning med to mindre utvalg med astma- og magesårpasienter. De finner i denne forbindelse en del interessante forskjeller. Statistisk pålitelig på 5%-nivået eller bedre er colitis-utvalgets høyere E-D-tendens enn astma-utvalgets, og astma-utvalgets høyere Im og H-D og lavere E-D enn magesårutvalgets. I forhold til Rosenzweigs norm-materiale viser for øvrig magesårutvalget minst, og astmautvalget størst uoverensstemmelse.

Seward og Morrison påpeker at en bør være oppmerksom på at magesårutvalget sannsynligvis omfatter to forskjellige personlighetstyper — en passiv type med overveiende vekt på impunitiv og behovspersistente reaksjoner, og en aktiv, men aggressiv og pågående type, med vekt på ekstrapunitiv og ego-defensive reaksjoner. Ved å behandle disse typer under ett vil mange av de karakteristiske forskjeller mer eller mindre dekkes over, og avviket fra normalmaterialet vil bli betydelig mindre enn det faktisk er.

At dette i høy grad er tilfelle, antyder deres GCR-materiale. Fra å være det utvalg som viser størst samsvar med normalmaterialet, rykker magesårutvalget nå ned som det som viser minst samsvar.

Faktisk viser det seg at magesårutvalget ligger lavere enn gjennomsnittet for Rosenzweigs nevrose- og psykose-utvalg. Resultatet reiser sterk tvil om holdbarheten av en påstand Rosenzweig har fremsatt, at GCR-verdier under 40 alltid vil være indikasjon på psykose.

Et karakteristisk fellestrekk ved alle de tre utvalg Seward og Morrison arbeider med, er et lite antall svar, lave ekstrapunitiv og lave ego-defensive tendenser. Resultatene antyder at denne kombinasjon av trekk gir uttrykk for en emosjonell tilbaketrekning og en disposisjon til å unngå personlig innvikling. En høy behovspersistens synes ikke alltid å gi uttrykk for god tilpassing. Er den behovspersistente kategori hovedsakelig ladet med impunitiv tendens og er den svært høy og oppnår den sin styrke vesentlig på bekostning av ego-defens, synes den mer å gi uttrykk for en overfladisk enn en virkelig god tilpassing, og mer for ønsketenkning enn konstruktiv problemløsning. Det foreligger her i det minste tolkningshypoteser som det er ønskelig blir tatt opp til nærmere undersøkelse.

En påfallende svakhet ved Seward og Morrisons undersøkelse er at den bygger på forholdsvis små utvalg. Denne svakhet gjelder også til en viss grad den undersøkelsen vi nå skal behandle.

Et område som er blitt relativt lite systematisk analysert fra test-psykologisk synsvinkel, er *selvmorderens* personlighetsstruktur. Selvmord er hos oss ikke noe stort sosialt problem, men heller ikke en helt ukjent foreteelse. I høyt industrialiserte land i Europa er det blitt påvist en selvmordshyppighet på 20, i enkelte tilfelle helt opp i 40, pr. 100.000 innbyggere.

Farberow har foretatt en undersøkelse med sikte på å bringe noe større klarhet over psykodynamiske sider ved den suicidal personlighet (19.) I tillegg til B-F-testen anvender han bl. a. *Minnesota Multiphasic Personality Inventory*, *Hildreth Feeling and Attitude Scale* og *Make a Picture Story Test*. Som fpp. nytter han mannlige pasienter ved et psykiatrisk hospital. Blant disse utgjør 64 en eksperimentalgruppe og resten — 32 — en kontrollgruppe. Eksperimentalgruppen består utelukkende av suicidaltruede pasienter som er blitt anbrakt i en særlig observasjonsavdeling. Kontrollgruppen er pasienter uten suicidal trekk, men for øvrig valt ut på en slik måte at den som helhet er relativt lik eksperimentalgruppen med hensyn til psykiatrisk diagnose, alder og utdannelse.

I Farberows eksperimentalgruppe er halvparten pasienter som faktisk har forsøkt selvmord, halvparten pasienter som har truet med selvmord. På tvers av denne inndeling er eksperimentalgruppen inndelt i to tallmessige like store grupper — halvparten er av psykiatere blitt vurdert som alvorlige, halvparten som ikke-alvorlige suicidaltruede. Tilsammen arbeider Farberow altså med 4 like store eksperimentelle utvalg, hvert bestående av 16 pasienter: 1) forsøkt selvmord, vurdert som alvorlige, 2) truet med selvmord, vurdert som alvorlige, 3) forsøkt selvmord, vurdert som ikke-alvorlige, og 4) truet med selvmord, vurdert som ikke-alvorlige.

Den anvendte dobbeltinndeling gir grunnlag for en meget allsidig statistisk behandling av det empiriske materiale. Ved hjelp av varians-analytiske metoder undersøkes hvorvidt det foreligger signifikante forskjeller mellom underutvalgene, og når dette er tilfelle, blir det så foretatt særskilte signifikansberegninger.

En sammenligning mellom Farberows data og Rosenzweigs normalmateriale gir en del interessante holdepunkter.

Rosenzweigs normalutvalg ligger til dels betydelig høyere enn alle Farberows utvalg med hensyn til ekstrapunitiv tendens, GCR, og S-E-mønstrene  $\underline{E}$  og  $\underline{E} + \underline{I}$ , mens Farberows utvalg i første rekke skiller seg ut ved høyere intro-punitiv tendens.

Når det gjelder forholdet mellom Farberows utvalg, finner vi konsistent høyere intro-punitiv og ego-defensiv tendens i eksperimentalutvalget, likeledes høyere GCR, høyere  $\underline{I}$  og  $\underline{E} + \underline{I}$ , men derimot konsistent lavere impunitiv og hindringsdominante tendens og lavere  $\underline{E}$ -faktor.

Farberows data viser at B-F-testen har evne til å differensiere mellom forskjellige pasientgrupper. Av særlig interesse er det å legge merke til at i forhold til både kontrollgruppen og gruppen med ikke-alvorlige tilfelle viser utvalget med alvorlige tilfelle seg å ha en betydelig sterkere intro-punitiv komponent. Videre synes denne komponent i relativt høy grad å være preget av rene I-svar.

Likledes er det av interesse å se at utvalget med alvorlige tilfelle ligge lavere enn utvalget med ikke-alvorlige når det gjelder ekstrapunitive tendenser, ikke minst fordi det ved bedømmelse av suicidalfare er svært alminnelig nettopp å legge vekt på grad og evne til ekstrapunitive reaksjoner.

Ved sammenligning av de forskjellige utvalgs ekstrapunitive tendenser fremkommer også en annen påfallende forskjell. Truselutvalget viser seg å ligge signifikant høyere enn forsøk-utvalget. Trusel om selvmord synes med andre ord ofte å være knyttet sammen med en viss ekstrapunitiv komponent.<sup>8</sup>

Et punkt Farberow trekker sterkt frem i diskusjonen av sine data, er at de ikke synes å peke hen på noen enhetlig, klart avgrenset suicidal personlighetstype. I forhold til alle enkeltkategorier finner han at eksperimentel- og kontrollgruppene går over i hverandre. Det som skiller den suicidal fra den ikke-suicidal pasient i det enkelte tilfelle, er oftest ikke ladningen på en enkelt faktor eller et enkelt trekk, men mer en mangel på balanse mellom et større antall forskjellige faktorer eller trekk.

Ved utelukkende å konsentrere seg om og sammenligne gjennomsnittsverdiene til de forskjellige gruppene på de enkelte kategoriene, slik Farberow gjør, blir en rekke forskjeller visket ut, og selve metoden gjør det umulig å analysere nærmere intraindividuelle forskjeller. Farberows undersøkelse er helt empirisk orientert. Han fremsetter ingen eksplisitte teorier eller hypoteser som han ønsker å etterprøve. Heller ikke foretar han noen systematisk sammenligning mellom resultatene på de forskjellige prøver han benytter seg av. Vi har ovenfor trukket frem bare enkelte sider ved Farberows materiale. Vi har hovedsakelig konsentrert oss om de resultater som antyder tolkningshypoteser som kan danne utgangspunkt for senere prøvinger.

## TESTENS DIFFERENSIERINGSEVNE OVERFOR GRUPPER MED FORSKJELLIG FRUSTRASJONSTOLERANSE

En undersøkelse av stor verdi når det gjelder å belyse et par av de antagelser B-F-testen er bygd på, er blitt foretatt av Hybl og Stagner (34). Hensikten med undersøkelsen var å finne frem til hvorvidt fru-

---

<sup>8</sup> Det ligger nær å tenke på den vanlige psykiatriske oppfatning at trusel om selvmord er en god indikasjon på en liten aktuell fare for selvmord. Farberow diskuterer dette synspunkt i sin avhandling og tar sterkt avstand fra det som et generelt prinsipp. Det avgjørende forhold er, mener han, ikke om trusler foreligger eller ikke, men hvilken form truslene har, om de omfatter en viss ekstrapunitiv komponent eller ikke.



strasjonstoleranse er en generell personlighetsfaktor, og hvorvidt psyko-terapi endrer et individs frustrasjonstoleranse.

Hybl og Stagner tar utgangspunkt i Rosenzweigs definisjon av frustrasjonstoleranse: «... et individs kapasitet når det gjelder å motstå en gitt frustrerende situasjon uten å fordreie livssituasjonens såkalte 'objektive' kjensgjerninger.» Og som operasjonelt kriterium på grad *fordreining* nytter de seg av «*nedsatt effektivitet* i behandling av en objektiv situasjon» (34, s. 163).

Et mål på frustrasjonstoleranse får de dermed ved å sammenligne fpp.s evne til oppgaveløsning umiddelbart før og etter frustrasjon. Til å fremkalle frustrasjon nytter de seg av høy klikke-lyd samtidig som de utsetter fpp. for skuffelser. Oppgavene som blir gitt, er av forskjellige typer. Ved å registrere ytelsesreduksjon på hver av oppgavetyperne og så korrelere resultatene finner de frem til samsvaret mellom prestasjonsforstyrrelser på forskjellige områder. Som fpp. nytter de 90 mannlige pasienter: 20 alkoholikere, 15 psykopater, 10 schizofrene, 25 nevrotikere og 20 ikke-psykiatriske tilfelle.

Hybl og Stagner konkluderer med at frustrasjonstoleranse, slik de har registrert denne egenskap, synes å være et generalisert aspekt ved individers personlighet. Normale og schizofrene viser høy frustrasjonstoleranse (liten reduksjon i effektivitet); nevrotikere, alkoholikere og psykopater vesentlig lavere. Mellom nevrotikere, alkoholikere og psykopater er det ingen signifikant forskjell, mens disse på den annen side er signifikant forskjellige fra de to andre grupper.

Det neste problem Hybl og Stagner stiller seg, er hvorledes to ekstremgrupper med høy og lav frustrasjonstoleranse vil skåre på Rosenzweigs B-F-test. Gruppen med høyest frustrasjonstoleranse (minst forstyrrelse på grunn av frustrasjon) består av 23 personer: 9 schizofrene og 14 normale, og gruppen med lavest av 22 personer: 2 psykopater, 7 alkoholikere og 13 nevrotikere. Foruten å regne ut gjennomsnittsverdier og forskjellene mellom dem for testens 6 skårings-kategorier, foretar de også en slik beregning for hver enkelt skåringsfaktor.

De finner i alt 8 signifikante forskjeller. Ekstremgruppen med høy toleranse oppnår signifikant flere B-P-, Im-, M-, e- og m-svar, gruppen med lav toleranse signifikant flere In-, M'- og I-svar. Størst differensierende evne har m-faktoren med  $t = 4.38$ , dernest følger M'-faktoren med  $t = 3.16$  og e-faktoren med  $t = 2.98$ . Først på fjerdeplass kommer så den mest differensierende kategori, In, med  $t = 2.92$ .

Den lave toleransegruppes sterkere intropunitive tendenser er av stor interesse sett i forhold til Frys undersøkelse over forskjellene mellom fanger og studenter. Som en vil huske, skilte fengselsklientelet i denne undersøkelse seg nettopp merkbart ut ved sine høyere intropunitive tendenser. Fengselsfanger synes altså å ha en sterk ladning av nettopp den faktor som særpreger nevrotikere, psykopater og alkoholister. Den mest differensierende reaksjonsform i Frys undersøkelse var for øvrig behovspersistens. Det er interessant å merke seg at dette også er den form som differensierer best i Hybl og Stagners undersøkelse.

For å undersøke hvorvidt psykoterapi har noen effekt i forhold til frustrasjonstoleranse, konsentrerer Hybl og Stagner seg om to grupper fpp., én gruppe bestående av 20 pasienter, alle med minst 10 timer individuell psykoterapi ( $M = 28,2$  timer), og en kontrollgruppe, også på 20 pasienter, men alle med under 10 timers behandling. Kontrollgruppen ble for øvrig valt ut slik, blir det opplyst, at den så vidt mulig ble lik eksperimentalgruppen både med hensyn til alder, utdanning og diagnose.

Ved sammenligning av de to grupperes gjennomsnittlige endring i toleranse blir det påvist en noe større endring i retning av større toleranse i eksperimentalgruppen. Likeledes blir det påvist en ikke ubetydelig overensstemmelse mellom toleranseendring og terapeutvurdering med hensyn til klinisk bedring.

Hybl og Stagner konkluderer med at «terapi har tendens til å øke frustrasjonstoleransen, skjønt forskjellene ikke er store». En må imidlertid være oppmerksom på, skriver de, at antall timer var lite og antall fpp. få. «Tendensen i tre av fire oppgaver er henimot større toleranse etter terapi . . . . Det er vår tolkning at terapi har virket til å redusere oppfatningen av trusel i nederlagssituasjoner, og/eller forøket persons selvutvurdering slik at han blir mindre kraftig forstyrret av det eksperimentelle nederlag.»<sup>9</sup> (34, s. 169).

Dessverre ble det ikke i Hybl og Stagners terapiundersøkelse foretatt noen før- og etterundersøkelse med B-F-testen.

Ved vurdering av deres undersøkelse må en særlig trekke frem den spesielle betydning som de tillegger frustrasjonstoleranse. Den operasjonelle betydning de gir begrepet, medfører at frustrasjonstoleranse nærmest blir lik distraksjonstoleranse.

Et meget viktig aspekt ved frustrasjonstoleranse fra en psykodynamisk synsvinkel er motivasjonsmessige forhold. En og samme grad av blokkering eller distraksjon vil vanligvis bety svært forskjellige ting for et individ alt etter intensiteten av det motiv som blir distraheret eller blokkert. Er motivets intensitet liten, vil en sterk distraksjon vanligvis ikke føre til noen opplevelse av trusel eller frustrasjon — endog om den generelle frustrasjonstoleranse er svært lav. Vi vil i slike tilfelle få en tilsynelatende toleranse, en toleranse som imidlertid ved nærmere studium vil vise seg å avspeile en motivasjonsmessig avstengning, en slags atolerant uanfektelse.

Det er høyst sannsynlig at det rent motivasjonsmessige grunnlag for oppgaveløsning var forskjellig blant Hybl og Stagners forsøksper-

---

<sup>9</sup> Denne tolkning er i samsvar med nyere forskningsresultater over terapi-prosessen — at det finner sted en endring i oppfatningen av truende situasjoner og av eget selv. Rosenzweig har et sted foreslått at psykoterapi nettopp blir sett på som en frustrasjons-toleranse-opppbyggende prosess. Hans utgangspunkt er psykoanalytisk: «opplevelse av små tolererbare doser av frustrasjon inntil motstand gradvis blir bygd opp» (64, s. 387). Mange terapeuter tar avstand fra denne såkalte immuniseringsteori, men i betraktning av terapi som toleranseoppbygning vil de fleste være enige.

soner. En sterk reduksjon i prestasjonsnivå vil allikevel kunne tas som uttrykk for liten frustrasjonstoleranse, men en mangel på reduksjon vil ikke alltid kunne tolkes som uttrykk for høyt toleransenivå.

Som tidligere nevnt ble E-D-kategorien på B-F-testen opprinnelig antatt å gi uttrykk for frustrasjonstoleranse. Tilsynelatende blir denne antagelse sterkt motsagt i første del av Hybl og Stagners undersøkelse. Noen signifikant forskjell mellom de to ekstremgrupper blir ikke påvist. Dette kan muligens skyldes at utvalgskriteriene, høy resp. lav distraksjonstoleranse, har liten differensieringsevne i forhold til frustrasjonstoleranse — idet begge deler gir uttrykk for liten toleranse. Vi skulle i så fall vente at begge ekstremgrupper var preget av en høy E-D-skåre, men for øvrig hadde forskjellig ladning på de forskjellige E-D-faktorer. Vi skulle vente at gruppen med høy distraksjonstoleranse fortrinnsvis var preget av en mer passiv holdning og hadde E-D-svar i første rekke av impunitiv karakter, og at gruppen med lav toleranse hovedsakelig hadde ekstrapunitive og intropunitive E-D-svar. En nærmere gransking av Hybl og Stagners data viser at dette er tilfelle. En analyse av de to ekstremgruppers sammensetning kaster ytterligere lys over forholdet. Den intolerante gruppen består av pasienter diagnostisert som psykopater, alkoholister og nevrotikere, den tolerante gruppen av 9 schizofrene og 14 ikke-psykiatriske tilfelle. Det vil si at av et samlet utvalg på 10 schizofrene viser hele 9 seg å tilhøre gruppen på 23 med høy distraksjonstoleranse. Å se på disse som preget av høy frustrasjonstoleranse vil si å motsi en vanlig klinisk betraktningssmåte. At den høye toleransegruppe også på annen måte er påvirket av dens schizofrene medlemmer, antyder de høyere ekstrapunitive og impunitive tendenser. Også disse reaksjonstyper er fra annet hold blitt påvist assosiert med schizofreni (2, 88).

En vesentlig konsekvens av vårt resonnement skulle være at begge Hybl og Stagners grupper var preget av relativt høye ego-defensive tilbøyeligheter. Hvis vi foretar en sammenligning med Taylors testnormer for menn (kfr. s. 93), viser dette seg å være tilfelle. Hvis vi dernest sammenligner med Rosenzweigs normmateriale, blir det hele mer usikkert. Rosenzweigs E-D-norm avviker ikke påfallende fra ekstremgruppenes gjennomsnittskårer. Om det er Rosenzweigs eller Taylors normer som i dette tilfelle danner det mest adekvate sammenligningsgrunnlag, er det vanskelig å avgjøre. Hybl og Stagners resultater kan derfor ikke sies hverken direkte å verifisere eller svekke Rosenzweigs hypotese at ego-defensive tilbøyeligheter på B-F-testen gir uttrykk for liten frustrasjonstoleranse.

I senere arbeider synes Rosenzweig å ha gått bort fra sin opprinnelige tolkningshypotese i forbindelse med ego-defens-kategorien.<sup>10</sup> Han synes etter hvert å være tilbøyelig til å betrakte så vel positive som negative avvik som tegn på liten frustrasjonstoleranse — samtidig som han ikke lenger begrenser seg til ego-defens-kategorien, men ser på alle markerte avvik på samme måte.

En naturlig implikasjon av dette synspunkt er utviklingen av en

---

<sup>10</sup> Dette kommer klart til uttrykk i Rosenzweigs tolkninger av Seward og Morrisons data (kfr. s. 45). Disse data reiser for øvrig sterkt tvil om holdbarheten av at lav E-D-tendens gir uttrykk for liten frustrasjonstoleranse.

«total-avvik-skåre», en oppsummering av et individs forskjellige skåre-kategoriers prosentvise avvik fra kategoriernes median eller gjennomsnittsverdier.

I en undersøkelse av Malamud blir en slik total-avvik-skåre nyttet som mål på frustrasjonstoleranse (72, s. 114). Malamud tar sikte på å undersøke effekten av gruppeterapi på individers frustrasjonstoleranse. Hun arbeider med to utvalg, ett med og ett uten gruppeterapeutisk erfaring. Hvert av utvalgene, som er forsøkt sammensatt på en slik måte at de er mest mulig like, omfatter 30 personer. I utvalget med terapi finner hun en signifikant endring i retning av mindre total-avvik-skårer, mens hun i utvalget uten terapi over det samme tidsrom ikke finner noen påvisbare endringer.

Dessverre er de data som foreligger om Malamuds undersøkelse, svært sparsomme, slik at noen inngående diskusjon av resultatene er umulig. Bruk av total-avvik-skårer til vurdering av frustrasjonstoleranse bygger imidlertid rent prinsipielt på et meget tvilsomt grunnlag. Det forutsetter at flertallet i det normalutvalg en arbeider med, har en optimal frustrasjonstoleranse. I vår kulturform vil dette sannsynligvis ytterst sjelden være tilfelle. Konformitet med flertallsnormer når det gjelder reaksjon på frustrasjon, vil oftest være en heller dårlig indikasjon på psykisk sunnhet, og som målsetting for psykoterapi et i høyeste grad inadekvat kriterium.

På mange måter virker bruk av tendenser som mål på frustrasjonstoleranse mer lovende enn bruk av total-avvik-skårer. Det er ennå ikke blitt publisert noen undersøkelser som klart påviser tendensberegningers verdi i forhold til testutgaven for voksne. Det åpner seg her imidlertid sentrale forskningsoppgaver, sett på bakgrunn av de meget løfterike resultater som er blitt påvist for barneutgaven av testen. Det er heller ikke utenkelig at utviklingen av forskjellige nye former for sekvensanalyser kan gi et bedre grunnlag for registrering av frustrasjonstoleranse enn det som er mulig ved hjelp av den tendensberegningens metode som hittil er foreslått.

## TESTENS DIFFERENSIERINGSEVNE OVERFOR GRUPPER MED FORSKJELLIG SOSIAL STATUS

De undersøkelser som er blitt behandlet tidligere i dette kapitel, viser at B-F-testen har en ganske bra differensieringsevne. Både ved sammenligning av sosiale og kliniske grupper er det blitt påvist signifikante forskjeller.

Vi vil til avslutning ganske kort trekke frem resultatene fra en nylig publisert undersøkelse av Lindzey og Goldwyn som tok sikte på å belyse B-F-testens evne til å differensiere mellom personer med høy og lav sosial status (42). Som fpp. brukte de 38 kvinnelige «college» studenter og som mål på sosial status to sosiometriske indekser.

Det er to hypoteser Lindzey og Goldwyn i første rekke tar sikte på å undersøke. De anser det som meget sannsynlig at et individs

aggresjonsretning alltid vil være medbestemmende for hvor godt likt individet blir blant sine medmennesker. Individer som blir meget godt likt, vil — mener de — oftest være preget av liten utoverrettet aggressivitet. Det vil først og fremst blant disse være impunitiv og intro-punitiv tendenser som er dominerende. Deres annen antagelse er at personer som er godt likt, også vanligvis vil ha en god sosial tilpassing. Ut fra dette mener de derfor en skulle vente at personer med høyest sosial status også vil skåre høyest med hensyn til GCR.

Når det gjelder den statistiske analyse, tar Lindzey og Goldwyn i bruk så vel korrelasjonsberegninger som sammenlikningsberegninger mellom ekstremgrupper. De finner i begge tilfelle et signifikant negativt samsvar mellom sosial status og ekstrapunitiv tendens. Personer med ekstremt høy sosial status synes etter resultatene å dømme i første rekke å være preget av impunitiv tendens. Også med hensyn til intro-punitiv tendens ligger «høy statusgruppe» over, men forskjellen er her mindre. I forbindelse med GCR får de resultater som går i motsatt retning av hva de ventet. Ved sammenligning av ekstremgrupper finner de faktisk — ved bruk av den ene av sine statusindekser — at «høy statusgruppe» skårer signifikant lavere enn «lav statusgruppe». Å bli ekstremt godt likt synes med andre ord ofte langt fra å være noe tegn på god sosial tilpassing.

Det er ytterst vanskelig å vurdere B-F-testen i lys av Lindzey og Goldwyns resultater. En meget bemerkelsesverdig side ved deres undersøkelse er at B-F-testen ble administrert omkring ett år etter innsamlingen av det sosiometriske materiale. At B-F-testen til tross for dette allikevel viser seg i stand til å differensiere, antyder i det minste en viss evne til å registrere konsistente personlighetstrekk.

## Kapitel VIII

### Testens validitet: 3. Samsvar med uavhengige kriterier

Et fellestrekk ved de undersøkelser som skal behandles i dette kapitel, er at de alle forsøker å belyse B-F-testens validitet ved å analysere samsvaret mellom testresultater og uavhengige kriterier. Det er blitt anvendt forskjellige typer av slike kriterier. Vi har funnet det hensiktsmessig å foreta en tredeling — bestemt av om de uavhengige kriterier representerer andre tester, atferdsobservasjoner eller eksperimentelle betingelser.

Bruk av eksperimentelle betingelser representerer en spesiell form for validitetsundersøkelse. Denne går ut på at en sammenligner testresultatene til de samme personer ved forskjellige, kontrollerte betingelser. Slike sammenligninger over tid vil ofte i forbindelse med personlighetsprøver kunne gi verdifulle opplysninger om prøvenes validitet.

#### SAMSVAR MED ANDRE TESTER

Vi har tidligere (s. 44) referert til en undersøkelse av Karlin og Schwartz over forholdet mellom B-F-testen og en generell intelligens-test. I samme undersøkelse ble det også foretatt en beregning av samsvaret med de forskjellige dimensjonene på en sosial intelligens-test (36).

To av de 5 dimensjonene på denne testen viste signifikant samsvar med B-F-testen: «Recognition of Mental State of Speaker» ga et negativt samsvar med ekstrapunitiv, og et positivt samsvar med intro-punitiv og impunitiv tendenser, og «hukommelse av navn og ansikter» ga et positivt samsvar med benektelser og intro-punitiv tendenser.

Et individ med høy Ex-skåre, antar Karlin og Schwartz, er vanligvis utoverrettet aggressivt og lite tilbøyelig til å modifisere sine reaksjoner med henblikk på andres behov. Individet vil vanligvis være preget av liten sosial sensibilitet og oppmerksomhet overfor andre. fra dette, mener de, kan man lett forklare det negative samsvar mellom Evne til å forstå andre.

Som forklaring på det sistnevnte samsvar fremsetter de den hypotese at benektelser og intro-punitiv tendenser gir uttrykk for en generell jeg-usikkerhet. En usikker person vil ha vedvarende behov for å fastslå sin egen posisjon i sosiale situasjoner. Personen blir sterkt opptatt med å observere andre, og utvikler lett en høy grad av sensibilitet overfor andres ansiktsuttrykk.

Begge deres hypoteser er post facto forklaringer. I et senere arbeid tar de utgangspunkt i det antatte forhold mellom svar på B-F-testen og oppfatningen av andres mentale tilstand. I stedet for å nytte en sosial intelligens-test foretar de en modifisering av B-F-testen, idet de anmoder fpp. ikke om å svare på vegne av en annen person, men om å angi hvilken sinnsstemning som preger den talende person på bildene. Ut fra statistiske betraktninger finner de frem til den passende eller normative oppfatning av hvert bilde i testen. Senere tar de B-F-testen i bruk på vanlig måte og korrelerer resultatene. Høyest korrelasjon finner de i forbindelse med B-P-kategorien, og de konkluderer med at under ellers like forhold synes personen med høy B-P-skåre «å være godt tilpasset, idet han opprettholder en realistisk kontakt med en situasjons krav» (86, s. 133). Den primære orientering som antydes ved B-P, synes å være konstruktiv og problemløsende. Schwartz og Karlin påpeker at den modifisering de har foretatt av B-F-testen, ennå ikke er helt tilfredsstillende standardisert. Selve tilnærmelsen åpner imidlertid interessante perspektiver. Ved å undersøke oppfatningen av personene som det reageres på, berører de noe av grunnlaget for svartendenser på testen.

Forsøk på å klargjøre den empiriske mening av testens forskjellige skåringskategorier kan ta mange veier. Schwartz (85) har f.eks. korrelert B-F-testen med Rorschach og et interesseinventorium. Hovedformålet med undersøkelsen var «å bestemme graden og arten av korrelasjon mellom skåringskategorier på projektive prøver og aktivitetspreferanser for derved direkte å undersøke de meninger som 'konvensjonelt' ble tillagt projeksjonstest-kategorier» (85, s. 141).

Til måling av aktivitetspreferanser nytter Schwartz *Sheviakov and Friedberg's Interest Inventory*. Dette interesseinventorium består av hele 600 ledd. Hvert ledd er en kort beskrivelse av en aktivitet som fpp. blir bedt om å angi om de liker, misliker eller stiller seg likegyldig overfor. Inventoriet inneholder eksempler på aktiviteter fra skole, hjem, sosiale, abstrakte og personlige forhold. Tilsammen gir det utgangspunkt for skåring i forhold til 41 kategorier.

Flere av B-F-testens kategorier viser seg å stå i et signifikant forhold til aktivitets- eller interessekategorier: Ekstrapunitiv tendenser f. eks. viser et positivt samsvar med 'interesse for skoleaktiviteter', 'aggresjon', 'ikke-identifisering med andre' og 'identifisering med andre', og et negativt samsvar med 'interesse for biologi' og 'interesse for matematikk'. Intro-punitiv tendenser viser et negativt samsvar med 'aggresjon', og et positivt samsvar med 'akseptering av autoriteter' og 'aversjon mot å påta seg dominerende roller'. Behovspersistente tendenser viser bl. a. et negativt samsvar med 'selvakseptering' og 'kritiske reaksjoner'.

sjoner'. Ego-defensive tendenser har et positivt samsvar med 'ikke-identifisering med andre' og et negativt samsvar med 'interesse for manipulering'. Og benektelser har bl.a. et positivt samsvar med 'opp-tatthet av renslighet', 'sport', 'skoleaktiviteter', 'musikk' og av 'streng-het overfor eget selv'.

«Rosenzweigs B-F-test er tilknyttet interesseinventoriet på en måte som antyder at 'aggresjonsretning' er signifikant korrelert med bestemte interesser,» konkluderer Schwartz. Han fortsetter: «Interesseinventoriets objektive data skaffer til veie et objektivt middel til å fastslå meningen av B-F-testens kategorier.» (85, s. 177).

Det er meget vanskelig å forstå Schwartz på dette punkt. Interesseinventoriet kan ikke uten videre sies å være noen mer objektiv metode enn B-F-testen, og dets kategoribetegnelser synes heller ikke å peke hen på noen mer avgrensede og meningsfylte variabler.

Schwartz's rent empiriske tilnærmelser gir en rekke signifikante korrelasjoner, men deres psykologiske mening er langt fra klar. Noen validitetsundersøkelse representerer tilnærmelsen strengt tatt ikke. Det samme kan for så vidt sies om Karlin og Schwartz's undersøkelser.

Heller ikke beregninger av korrelasjoner mellom B-F-testen og Rorschach gir mer enn antydninger om de to metoders forhold til hverandre. Ved interkorrelasjon av 31 Rorschach-variabler og 8 variabler på B-F-testen finner Schwartz 28 samsvar med en statistisk pålitelighet på 5%-nivået eller bedre. Bl. a. finner han et negativt samsvar mellom S% og In- og Im-kategoriene og et positivt samsvar mellom  $\Sigma C$  og ego-defens-kategorien.

Mens begge tester hver for seg viser seg å være nær tilknyttet interesseinventoriet, er de fleste av interkorrelasjonene mellom dem insignifikante og dessuten ytterst vanskelig å tolke. «Dette kan kanskje skyldes den ting at de to instrumenter reflekterer forskjellige reaksjonsnivåer . . . .» skriver Schwartz. «De synes å måle forskjellige aspekter ved interesser . . . .» (85, s. 177).

En undersøkelse av Fisher og Hinds (20) om aggresjonskontrollen hos forskjellige personlighetsstrukturer inneholder en del data som belyser forholdet mellom B-F-testen og en del andre tester, bl.a. også Rorschach. En ulempe ved undersøkelsen er at det utelukkende blir gjort bruk av 12 av bildene på B-F-testen, at skåringen hele tiden bare tar sikte på å klargjøre reaksjonsretning, og at utvalget av fpp. er svært lite. Det er av denne grunn umulig å trekke noen sikre slutninger av resultatene, men de gir mange interessante antydninger. Vi vil her bare konsentrere oss om noen av korrelasjonene som blir beregnet for deres normalutvalg av fpp. Dette utvalg består av 11 menn og 15 kvinner, med en gjennomsnittsalder på litt over 23 år. Vi vil senere komme tilbake til andre sider ved undersøkelsen. Tabell 4 gir en oversikt over korrelasjonen mellom to av retningskategoriene på den forkortede B-F-test og «aggressive tendenser» ifølge forskjellige andre tester.

Høyest korrelasjon finner altså Fisher og Hinds i forbindelse med Rorschach og *Feeling Questionnaire*, et spørreskjema som tar sikte på



TABELL 4

*Korrelasjonen mellom B-F-testen (forkortet) og andre tester*

| Andre tester                         | Ex    | In    |
|--------------------------------------|-------|-------|
| Feeling Questionnaire . . . . .      | .42   | — .36 |
| Hostile Word Associations . . . . .  | — .06 | .19   |
| TAT (åpen aggresjon) . . . . .       | .01   | .01   |
| TAT (indirekte aggresjon) . . . . .  | .35   | — .20 |
| Rorschach (åpen aggresjon) . . . . . | .37   | — .44 |

å gi uttrykk for i hvilken grad fpp. betraktet seg selv som aggressive. Lavest korrelasjon viser seg å være til stede med åpen aggresjon på TAT og aggressive svar på en ordassosiasjonsprøve. Også i en annen undersøkelse (42) er det blitt påvist ikke å foreligge noe samsvar mellom B-F-testen og en slik prøve. Det er imidlertid kanskje i første rekke bør merke seg ved resultatene, er den markerte forskjell i samsvaret med åpen aggressivitet på TAT og indirekte, rasjonalisert aggressivitet på denne prøve. Forskjellen antyder at aggressivitet på TAT henspiller på forskjellige ting alt etter hvilke tolkningsprinsipper som blir anvendt.

Forholdet mellom B-F-testen og TAT er av flere blitt tatt opp til nærmere undersøkelse.

En av de første opplysninger Rosenzweig gir om B-F-testens validitet, er faktisk at ekstrapunitiv skåre på testen er blitt funnet å gi en positiv korrelasjon på +.74 med ekstrapunitive tilbøyeligheter vurdert uavhengig på grunnlag av TAT (65, s. 22).

I skarp kontrast til dette resultat står resultatene av en senere undersøkelse av Lindzey (40). Lindzey konsentrerer seg om bare 4 av TAT-bildene og foretar en beregning av ekstrapunitive tendenser ved en oppsummering av antall tilfelle hvor «selv-figuren» legger for dagen aggressive handlinger mot andre figurer, og intro-punitive tendenser ved å oppsummere antall tilfelle av aggressive handlinger «selv-figuren» retter mot seg selv.

«Hverken ekstrapunitivitet eller intro-punitivitet slik de blir målt ved B-F-testen,» konkluderer Lindzey, «viser noen korrelasjon med de samme dimensjoner slik disse blir målt med TAT» (40, s. 320).

Noen nærmere sammenligning mellom Rosenzweigs og Lindzeys undersøkelse er vanskelig, da Rosenzweig hverken gir opplysninger om sitt utvalgs størrelse eller på hvilken måte han har anvendt TAT. Det ligger imidlertid nær, i lys av Fisher og Hinds's resultater, å anta at Rosenzweig i motsetning til Lindzey har lagt hovedvekten på indirekte og rasjonalisert aggressivitet i skåringen av TAT-protokollene.

Denne forklaring blir ytterligere støttet av resultatene fra en undersøkelse av Tejessy (42). Undersøkelsen tok sikte på å beregne korrelasjonen mellom B-F-testen og hver enkelt av 10 forskjellige TAT-tegn på aggressivitet. Trekker vi imidlertid også inn i bildet resultatene av en undersøkelse av Lindsey og Goldwyn, hvor det også blir beregnet korrelasjoner med forskjellige TAT-tegn, blir situasjonen ytterst uklar. Tabell 5 gir en oversikt over noen av korrelasjonene som ble funnet i de to undersøkelser.

TABELL 5

*Korrelasjonen mellom B-F-testen  
og forskjellige aggresjonstegn på TAT\**

| TAT-tegn   | Tejessys unders. (N = 20) |      | Lindzey og Goldwyns unders. (N = 38) |      |
|--|---------------------------|------|--------------------------------------|------|
|  | Ex                        | In   | Ex                                   | In   |
| Aggressive vendinger . . . . .                         | .49                       | -.46 | -.09                                 | -.04 |
| Død, resultat av ytre krefter                          | .20                       | -.08 | -.11                                 | -.08 |
| Vold . . . . .   | .40                       | -.39 |                                      |      |
| Død, nederlag etc. for ikke-<br>„selv-figur” . . . . . | .30                       | -.19 |                                      |      |
| Feilgjenkjenning . . . . .                             | .14                       | -.03 | .29                                  | -.25 |
| Sterke aggressive fantasier                            | .33                       | -.26 | -.17                                 | .08  |
| Språk . . . . .  | .38                       | -.41 |                                      |      |
| Generell taksering av<br>aggressivitet . . . . .       |                           |      | -.12                                 | .19  |

\* For å oppnå en statistisk pålitelighet på 5%-nivået kreves i Tejessys undersøkelse en korrelasjonskoeffisient på .38, mens kravet til en slik pålitelighet i Lindsey og Goldwyns undersøkelse er .32.

Mens alle Tejessys enkeltresultater går i ventet retning og hele 6 av korrelasjonene oppnår en viss statistisk pålitelighet, antyder Lindsey og Goldwyns resultater en praktisk talt fullstendig mangel på korrelasjon mellom de to prøver. Muligheten for at forskjellen avspeiler forskjell i bruk av tolkningsprinsipper er ytterst liten. Derimot synes administrasjonsformen å være noe forskjellig, idet Lindsey og Goldwyn nytter TAT som gruppetest. Hvor stor betydning som bør tillegges dette, er det ytterst vanskelig å ha noen sikker mening om. Heller ikke i hvilken utstrekning utvalget med fpp. i de to undersøkelser var like — psykologisk sett.

Bruk av resultater på andre prøver som validitetskriterium reiser, som en vil forstå, meget kompliserte problemer. Grunnen er i første rekke den at «andre prøver» vanligvis ikke har noen tilfredsstillende klarlagt validitet. Å gjøre bruk av slike prøver vil derfor lett bli en fullstendig famling i mørket, en situasjon hvor en blind skal lede en blind.

## SAMSVAR MED MANIFEST ATFERD

Brun-Gulbrandsen gjør i sin skolehjemsundersøkelse bl.a. greie for anvendelsen av Rosenzweigs B-F-test blant skolehjems elever og kommer også inn på forholdet mellom E-faktoren i testen og tilbøyelighet til å komme i konflikt med omgivelsene, til å vise uønsket atferd, slik dette gir seg uttrykk i lærervurderinger.

Han finner ingen nevneverdig korrelasjon mellom disse variabler ( $r = + .08$ ), men trekker samtidig frem at lærervurderinger er et heller tvilsomt kriterium: «. . den lave korrelasjon svekker ikke hypotesen definitivt» (10, s. 184).

Når det gjelder vurderinger av atferdsvariabler, er det flere hensyn en må være oppmerksom på. Svært ofte vil vurderingsstandardene være forskjellig fra en person til en annen. Ofte vil også referanserammen være en annen enn den en erfaren psykolog i første rekke har for øye når han sikter til «konflikt med miljøet», «ekstrapunitivitet» etc. Det er derfor med stor interesse en legger merke til følgende uttalelse i den omtalte undersøkelse: «Ved inspeksjon av skårene for E-faktoren og ved sammenligning med det inntrykk vi har av hver enkelt fp. etter 3—6 måneders samvær med samtlige fp., har vi et absolutt inntrykk av at det er en ikke ubetydelig overensstemmelse mellom de nevnte skårer og tendenser til aggressiv atferd. Noe eksakt uttrykk for dette har vi imidlertid ikke funnet.» (10, s. 184).

Brun-Gulbrandsens behandling gir ingen sikre holdepunkter når det gjelder testens validitet. Muligheter for «opinion-errors» er i sterk grad til stede.

Det samme kan vanskelig sies om en undersøkelse av Sinaiko (90). Det ble her tatt særlige forholdsregler nettopp med sikte på å få mest mulig ensartede vurderinger, samtidig som testen ble skåret helt uavhengig av atferdsvurderingene.

Formålet med Sinaikos undersøkelse var å finne frem til en hensiktsmessig utvalgsmetode til bruk ved ansettelse av avdelingssjefer i et større varemagasin. Som fpp. ble nyttet 53 av firmaets 58 avdelings-sjefer. Først ble hver enkelt vurdert med hensyn til arbeidsytelser. Dernest gjennomgikk de et testbatteri bestående av to intelligens-tester og B-F-testen.

Ved korrelasjonsberegninger mellom dyktighetsvurderingene og intelligens-test-resultatene finner Sinaiko ikke noe korrelasjon, derimot finner han flere signifikante korrelasjoner mellom B-F-testen og dyktighetsvurderingene og også en signifikant korrelasjon mellom denne testen og lengde av ansettelsestid.

Sinaikos resultater antyder at dyktighet som avdelingssjef er positivt korrelert med tilbøyelighet til intro-punitiv og behovspersistente reaksjoner (.28 resp. .38), negativt korrelert med ego-defensive og ekstrapunitiv tendenser (— .48 resp. — .31).

«Tar en for seg de krav som stilles en avdelingssjef,» skriver Sinaiko, «følger disse forhold et logisk mønster» (90, s. 39). Ledelsen krever at deres avdelingssjefer skal legge for dagen en vedvarende toleranse i kontakt med kunder. «Kunden har alltid rett» er en innstilling som ledelsen forsøker å oppmuntre. Firmaets politikk når det gjelder forholdet til kunder, har en sterk intro-punitiv komponent. Videre fremheves sterkt at en avdelingssjef bør legge for dagen en mest mulig avpasset, problemløsende innstilling, være minst mulig forsvarsberedt og minst mulig innstilt på å bebreide og anklage.

At de dyktigste avdelingssjefer viser en høyere grad av intro-punitiv tendenser, er derfor hva en skulle vente. Det er også den negative korrelasjon mellom dyktighet og ekstrapunitiv og ego-defensive tendenser.

Når det gjelder forholdet mellom skåringskategoriene og ansettelsestid, er det bare en signifikant korrelasjon. De lengst ansatte avdelingssjefer viser mer impunitiv tendenser enn de kortest ansatte. Også denne korrelasjon er forståelig.

I det store og hele viser B-F-testen seg i stand til å differensiere mellom de mest og de minst dyktige avdelingssjefene. Testen er imidlertid ikke noe utpreget godt utvalgsinstrument, og brukt alene vil den føre til usikre resultater. Det er av betydning å merke seg i denne forbindelse at evnetestene var uten enhver differensieringsevne. «Undersøkelsen antyder,» konkluderer Sinaiko, «at Rosenzweigs B-F-test måler faktorer som er knyttet sammen med yrkesmessig suksess som avdelingssjef, og at den kan være av verdi i et personelluttags-program» (90, s. 42).

Testen synes ifølge Sinaikos resultater å gi grunnlag for forutsigelse om vanlig, dagligdags atferd. Et spørsmål som øyeblikkelig reiser seg, er imidlertid hvorledes dyktighetskriteriet i Sinaikos undersøkelse var konstruert. Var det et reliabelt kriterium? Var det valid i den forstand at det virkelig ga uttrykk for dyktighet — for nettopp *den type dyktighet* som vi ovenfor har siktet til?

Sinaiko opplyser at han som kriterium på dyktighet nyttet et kvantitativt mål, basert på vurderingsskalaer (Executive Personell History Form) for 7 grunnleggende egenskaper: karakter, intelligens, intuisjon, erfaring, tilpassing, spesielle ferdigheter og utvikling. Hver avdelingssjef (section manager) ble vurdert av en av firmaets fire inspektører (floor superintendents). Alle vurderinger ble deretter kontrollert og samordnet av lederen for personalkontoret for å få mest mulig ensartede vurderinger.

Alle avdelingssjefer som undersøkelsen omfatter, hadde vært ansatt i minst 3 måneder. Inspektørene som foretok atferdsvurderingene, var derfor godt kjent med avdelingssjefene.

Atferdsvurderinger kan under gunstige forhold gi et relativt objektivt materiale, men under andre forhold være høyst upålitelige. Noen opplysninger om inspektørenes trening og evner som observatører gir ikke Sinaiko, men det ligger likevel nær å tillegge undersøkelsen stor betydning når det gjelder å kaste lys over B-F-testens validitet.

Også i utarbeidelsen av et prøvebatteri for utvalg av politikonstabler har en funnet B-F-testen av verdi (17). Av interesse merker en seg at det i denne forbindelse er ego-defensive og intropunitiv tendenser som gir positiv korrelasjon med dyktighetskriteriet, en tjenestevurdering av en overordnet politioffiser etter 10 ukers praktisk tjeneste. I begge tilfelle er imidlertid korrelasjonen relativt beskjedne, og det er i første rekke det forhold at det blant et større antall andre tester som samtidig ble prøvd, overhodet ikke ble funnet noen korrelasjon med kriteriet, som får Du Bois og Watson til bl.a. å anvende B-F-testen i deres endelige utvalgsbatteri.

Du Bois og Watsons undersøkelser gir imidlertid svært lite når det gjelder å belyse B-F-testens validitet. Deres uavhengige kriterium — tjenestevurderinger — er sannsynligvis et høyst sammensatt kriterium, og noen systematiske forsøk på å oppnå ensartede vurderinger blir ikke gjort.

Både Sinaikos og Du Bois og Watsons undersøkelser belyser B-F-testens praktiske anvendbarhet. Også en undersøkelse av Schmeidler (84) påpeker muligheter i denne retning, uten at undersøkelsen ble foretatt nettopp med dette for øye.

Det har lenge vært et velkjent fenomen at personers innstilling til ekstrasensorisk persepsjon (ESP) står i et nært forhold til deres faktiske prestasjoner på dette område. Ikke bare er det blitt påvist at velvillig innstilte personer gjennomgående gjør det bedre ved clairvoyance-eksperimenter enn uinteresserte, men også at personer preget av pessimisme og negativisme ofte har en tendens til å skåre signifikant dårligere enn det en skulle vente ut fra sannsynlighetsberegninger. Likeledes har en påvist at personers generelle tilpassingsevne, fastslått på grunnlag av f.eks. Rorschach, har stor innvirkning på deres evne til ekstrasensorisk persepsjon. Personer som er tolerante og vennlige, har vist seg vanligvis å ha mer ESP-suksess enn personer med sterke aggressive tendenser.

Med bakgrunn i dette skulle en altså vente at personer med impunitiv tendens på B-F-testen skulle oppnå høyere ESP-skårer enn de hvis reaksjonstendenser i første rekke går i ekstrapunitiv retning.

Schmeidlers data omfatter materiale fra flere enkeltundersøkelser foretatt på forskjellige utvalg med fpp. Tilsammen består hennes totalmateriale av hele 495 studenter. Hennes resultater viser en klart positiv korrelasjon, signifikant på .003-nivået, mellom impunitiv tendens og ESP-skåre, og en negativ korrelasjon, signifikant på .005-nivået, mellom ESP-skåre og ekstrapunitiv tendens.

Når hun deler sitt materiale inn i personer preget av henholdsvis positiv og negativ innstilling til mulighetene for ESP, finner hun en påfallende forskjell. Utelukkende blant de positive er det et høyt signifikant samsvar mellom ESP-skåre og skåre på B-F-testen. Videre finner

hun ved sammenligning av de forskjellige utvalg av fpp. at det bare er i de tilfelle hvor ESP-eksperimentet ble introdusert på en slik måte at det sannsynligvis fremkalte frustrasjon, at det ble registrert signifikante korrelasjoner.

Schmeidler betrakter sine resultater som uttrykk for at det er nødvendig å kjenne til hvor frustrerende en eksperimentell ESP-situasjon vil oppleves før det er mulig å foreta noen forutsigelse av individers atferd i denne, ut fra deres svar på B-F-testen. Har en imidlertid opplysninger om fpp.s mer generelle mønstre av behov og innstillinger — og særlig kjennskap til om den eksperimentelle situasjon vil bli oppfattet som frustrerende, vil B-F-testen kunne være av stor verdi.

På grunnlag av dag-til-dag-observasjoner av studentene i sitt forsøksmateriale hevder hun å ha kommet frem til at en *høy* impunitiv skåre på B-F-testen kan stå for to forskjellige personlighetsmønstre. «Mange var vennlige, milde og hjelpsomme i deres daglige kontakter med klassekamerater, og samarbeidsvillige og nærmest umerkelige i klasseværelset . . Disse studentene fikk i alminnelighet høy ESP-skåre. Mens andre, innbefattet flere med et uvanlig høyt antall impunitive svar, virket hemmet og kalde. De ga inntrykk av å gjøre bruk av en forsiktig, selvbevisst selvkontroll, og de fikk i alminnelighet ESP-skårer på sannsynlighetsnivået eller lavere.» (84, s. 342).

Med bakgrunn i Rosenzweigs hypoteser om B-F-testens forskjellige kategorier skulle vi vente at de to personlighetstyper har forskjellige faktor-ladninger på den impunitive kategori. Dessverre foretar Schmeidler ikke noen analyse av forholdet mellom ESR-skåre og reaksjonsformtendenser. Hennes påpekning av at den impunitive kategori står for forskjellige personlighetstyper, har en på sett og vis tatt hensyn til ved selve testens konstruksjon. Korrelasjonsberegninger med reaksjonsform-kategoriene kunne ha kastet lys over sider ved testen som mange forskere svært ofte synes å ha en tendens til å overse. Schmeidlers påpekning av at flere av dem som virket hemmet og kalde, hadde et *uvanlig høyt* antall impunitive svar, støtter for øvrig indirekte Rosenzweigs hypotese — at alle former for markerte avvik kan tolkes som uttrykk for emosjonelle tilpassingsvansker.

Både Schmeidlers og Sinaikos resultater antyder at det foreligger samsvar mellom B-F-testen og manifest atferd. Et fellestrekk ved undersøkelsene er at det som fpp. begge steder gjøres bruk av et psykologisk normalmateriale. Det ligger nær å spørre om en tilsvarende korrelasjon også er til stede blant psykologisk mer anormale utvalg, blant f.eks. psykiatriske pasienter.

Det foreligger i dag et par undersøkelser som er blitt foretatt med sikte på å belyse nettopp forholdet mellom resultater på B-F-testen og pasienters atferd på psykiatriske hospitaler. I en undersøkelse av Albee og Goldman blir ulykkes- og skaderapporter tatt som utgangspunkt for klassifisering av pasienter som ekstrapunitive i den grad de var skyldige i å påføre andre skader, og som intropunitive i den grad de tilfeldig eller med hensikt påførte seg selv skader (3).

Klassifisering av pasienter på dette grunnlag var i en tidligere undersøkelse (2) blitt påvist å ha prognostisk verdi. Det hadde vist seg mulig ut fra kjennskap til hvorledes pasientene rettet sin aggressjon, å slutte noe om deres sannsynlige kliniske bedring. Intropunitiv pasienter viste seg gjennomgående å ha betydelig bedre prognose enn ekstrapunitiv — både generelt og spesielt, dvs. i de tilfelle hvor pasientenes diagnose for øvrig var den samme.

Hvis det skulle vise seg å foreligge en høy korrelasjon mellom Ex- og In-kategoriene på B-F-testen og klassifiseringen på grunnlag av ulykkes- og skaderapportene, ville dette følgelig antyde noe om testens prognostiske verdi.

Albee og Goldman finner at hele 70 % av de 65 pasientene de arbeider med, oppnår skåreverdier på den ekstrapunitiv og intropunitiv kategori som overgår tredje kvartil i Rosenzweigs opprinnelige normmateriale, men de finner ikke noen korrelasjon mellom testresultater og deres atferdsklassifisering. De foretar også en analyse av forholdet mellom testskårer og prognose. Mens de på ny finner et ganske godt samsvar med deres klassifisering, finner de imidlertid ikke noe direkte samsvar mellom Ex- og In-kategoriene på testen og klinisk bedring. Høyst indirekte mener de allikevel testen kan ha verdi; i det tilfelle hvor pasientene er ekstrapunitiv og deres testskårer er sterkt intropunitiv, viser prognosen seg å være svært dårlig.

En nærliggende innvending som kan rettes mot Albee og Goldmans undersøkelse, er at deres uavhengige mål på ytre atferd ikke er spesielt tilordnet vanlige reaksjonstendenser på frustrasjon, men i første rekke omfatter mer eksplosjonsartede og ukontrollerte atferdsmønstre, som kanskje er inkonsistente i forhold til pasientenes vanlige reaksjonstendenser. Et sentralt spørsmål blir derfor i hvilken grad det innenfor et psykiatrisk sykehusmiljø faktisk foreligger en høy grad av konsistens i pasientenes reaksjonstendenser.

Et indirekte svar på dette får vi gjennom resultatene av en undersøkelse av Simos (89). Simos's utgangspunkt er å forsøke å klargjøre korrelasjonen mellom B-F-testen og forskjellige uavhengige kriterier. Han arbeider med et utvalg på 50 pasienter ved et psykiatrisk sykehus.

I første rekke konsentrerer han seg om to atferdskriterier: observasjon av fpp.s atferd under en eksperimentell stress-situasjon, og uttalelser av psykiatere ved sykehuset om deres daglige atferd. I begge tilfelle blir fpp.s atferd vurdert i forhold til B-F-testens retnings- og formkategorier.

Korrelasjonen mellom dominerende reaksjonsretning og -form, vurdert på grunnlag av testresultater og stress-observasjoner, viser seg å være klart positiv. Den gjennomsnittlige statistiske pålitelighet ligger på noe under 10%-nivået, med et enkelt utslag ned på 1%-nivået.

Simons påpeker sterkt at resultatene er høyst usikre på grunn av små hyppigheter ved enkelte av kategoriene, men hevder på den annen side at den «tilnærmede signifikans av korrelasjonen mellom stress-

vurderinger og B-F-testen synes å gi endel støtte til verdien av testen som et ideografisk instrument . . . » (89, s. 330).

Av særlig interesse i den foreliggende sammenheng er at det overhodet ikke viser seg å være noen korrelasjon mellom atferdsvurderingene av psykiatere og resultatene på B-F-testen. Det ligger derfor ytterst nær å tolke Simos data som en påvisning av at det ikke foreligger noen høy grad av konsistens i pasientenes reaksjonstendenser. Dette blir for øvrig også illustrert mer direkte ved beregning av korrelasjonen mellom stress-vurderingene og psykiaternes vurderinger. Disse viser både med hensyn til reaksjonsretning og reaksjonsform en liten positiv korrelasjon, men en relativt mindre korrelasjon enn korrelasjonen mellom testresultater og stress-vurderinger.

Det er ikke urimelig å anta at psykiaterne på grunn av deres spesielle trening i uforholdsmessig grad har vært influert av mer eksplosive og voldsomme handlinger, slike som noteres i ulykkes- og skaderapporter, mens stress-vurderingene tross alt mer er assosiert med pasientenes vanlige, daglige atferd. En slik antagelse forklarer i det minste overensstemmelsen med Albee og Goldmans resultater.

En konsekvens av resonnementet ovenfor er at vi ser på B-F-testen som et instrument som hovedsakelig registrerer personers manifeste reaksjonstendenser overfor «mildt frustrerende, ofte forekommende situasjoner». At denne betraktningssmåte ikke er generelt holdbar, antyder imidlertid resultatene av en undersøkelse av Lindzey og Goldwyn (42). I sin form har denne undersøkelse mange likhetspunkter med Simos eksperimentelle metode. Også Lindzey og Goldwyn lar sine fpp. utsettes for en eksperimentell stress-situasjon, og også de foretar korrelasjonsberegninger mellom atferdstaksringer og Ex- og In-kategoriene på B-F-testen.

I motsetning til Simos finner Lindzey og Goldwyn imidlertid ikke noen signifikante korrelasjoner. Deres utvalg med forsøkspersoner er ytterst lite idet det bare omfatter 17 personer, og deres atferdsobservasjoner er sikkert heller ikke særlig pålitelige, idet stress-situasjonene ikke varte mer enn ca. 15 minutter, men til tross for slike iøynefallende metodiske svakheter er forskjellen med Simos resultater allikevel bemerkelsesverdig.

En meget påfallende forskjell ved undersøkelsene er bruken av fpp. Mens alle fpp. i Lindzey og Goldwyns utvalg er aktive studenter, konsentrerer Simos seg som nevnt utelukkende om psykiatriske pasienter. Det er ikke utenkelig at B-F-testen registrerer forskjellige sider ved personligheten hos forskjellige personer. Også i foregående avsnitt ble vi tvunget til å reise en slik problemstilling. Iallfall åpner det seg her en meget interessant problemstilling.

## EKSPERIMENTELLE UNDERSØKELSER

En undersøkelse over en tests validitet finner vanligvis sted ved at det blir foretatt korrelasjonsberegninger mellom testresultatene for en gruppe personer og et eller annet uavhengig validitetskriterium.



En mer spesiell form for validitetsundersøkelse er å sammenligne testresultatene for de samme personer ved forskjellige, eksperimentelt kontrollerte betingelser. Særlig i forbindelse med projektive metoder er dette en meget hensiktsmessig fremgangsmåte, da det er ytterst vanskelig å finne adekvate ytre validitetskriterier. Bruk av andre tester som kriterium forutsetter at en faktisk disponerer over valide tester. Dette er i dag ikke ubetinget tilfelle. Bruk av atferdsobservasjoner er heller ikke alltid vel egnet. Projektive prøver tar ofte sikte på å registrere latente disposisjoner eller trekk. Disse trekk har selvsagt atferdsmessig relevans, men de er ikke atferdsekvivalente — de lar seg ikke registrere direkte. At B-F-testen ikke alltid korrelerer med manifest atferd, vil derfor ikke nødvendigvis si at den er invalid.

Bruk av eksperimentelle metoder åpner spesielle muligheter. Det foreligger imidlertid i forbindelse med slike metoder også en rekke muligheter for grove feilslutninger. Eksperimentelle undersøkelser byr ofte på store vansker, og noen lettvinnt og ufeilbarlig valideringsmetode er det ikke.

Vi har tidligere behandlet en undersøkelse av B-F-testen basert på en eksperimentell metode. Vi sikter til Malamuds undersøkelse (kfr. s. 52), over forholdet mellom endringer i totalavvikskårer i to utvalg av fpp., det ene med, det andre uten psykotераpeutisk erfaring. Som nevnt er det vanskelig å foreta noen nærmere vurdering av Malamuds resultater, men selve den metodiske tilnærming fortjener all mulig oppmerksomhet.

To av de tre undersøkelser vi skal behandle i dette kapitel, er beheftet med avgjørende mangler. At vi allikevel har funnet det ønskelig å gi dem en relativt bred omtale, er bestemt av vårt håp om å gi illustrasjoner som kan tjene som veiledning for senere studier på dette området.

Vi vil ta utgangspunkt i en undersøkelse av Franklin og Brozek (21) med formål å finne frem til hvorledes forskjellige typer av diett virker på personer som i lengre tid har vært utsatt for underernæring.

Undersøkelsen omfatter tre faser: Først en standardiseringsfase på 3 måneder hvor fpp., i alt 36 menn i alderen 20 til 33 år, fikk rikelig med mat (3492 kalorier pr. dag). Dernest en 6 måneders periode med lite mat (1570 kalorier pr. dag), og til slutt en 3 måneders rehabiliteringsperiode hvor fpp. ble delt inn i 4 grupper med en gjennomsnittlig daglig ernæring på henholdsvis 2377, 2692, 3123 og 3392 kalorier.

B-F-testen ble administrert første gang etter 24 uker med underernæring, og annen gang etter 12 ukers ernæringsmessig rehabilitering.

Ved slutten av underernæringsperioden viste fpp. et gjennomsnittlig tap på 24 % av deres vekt. Også mer psykologiske funksjoner, blir det opplyst, ble blokkert eller forstyrret. «Frustrasjonsreaksjonene til våre fpp. i underernæringsperioden, sammenlignet med deres 'normale' frustrasjonsreaksjoner, viste klart en endring i intensitet, retning og type av reaksjoner . . . Fpp.s normale aktivitetsmønstre ble alvorlig svekket . . . Frustrasjonsterskelen og frustrasjonstoleransen sank på-

fallende. Fpp.s atferd fra dag til dag viste klart både generaliserte emosjonelle tilstander (f. eks. sinne, angst) og mer spesielle og målrettede tilstander eller «følelser» (f.eks. fiendtlighet, sjalusi, mindreverd) i overensstemmelse med Sargents (83) klassifikasjon av frustrasjonstilstander» (21, s. 297).

Etter 12 ukers rehabilitering hadde fpp. i de 4 kalorigruppene fra lavest til høyest, gjenvunnet gjennomsnittlig 21, 30, 41 og 57 % av sitt vekttap. «På dette tidspunkt,» skriver Franklin og Brozek, «var sult og matdeprivasjon ikke lenger et gruppeproblem . . . Den økede tilgang på mat førte ikke bare til en bedring og en tilsvarende tilbakevending til «normal» atferd, men også til at frustrerende påvirkninger og reaksjoner som var til stede under trykket av underernæring, ble borte eller drastisk redusert.» (21, s. 297).

I den grad B-F-testen registrerer hvorledes et individ møter frustrasjon, skulle en vente å finne forskjeller i reaksjonsmønstre ved de to betingelser. En skulle bl.a., dersom Franklin og Brozeks påstand er holdbar, vente å finne at frustrasjonstoleransen var mindre i underernæringsperioden.

Franklin og Brozek presenterer sine resultater utelukkende i form av gruppegjennomsnitt på svarkategoriene ved de to tidspunkter. Test-resultatene i underernæringsperioden viser seg *stort sett å være lik resultatene i rehabiliteringsperioden*. Det er ikke utenkelig at de ved å konsentrere seg om gjennomsnittsverdier dekker over vesentlige individuelle endringer, men mangelen på systematiske endringer er allikevel påfallende.

«Den foreliggende undersøkelse,» konkluderer de, «reiser tvil om validiteten av B-F-testen som et objektivt mål på frustrasjonsreaksjoner . . . Resultatene setter et spørsmålstegn ved testens validitet. Inntil testen mer adekvat møter kravene til et objektivt instrument, bør den ikke bli brukt i rutinemessig diagnostisk eller klinisk arbeid, men bli begrenset til utprøvende og eksperimentell anvendelse.» (21, s. 300).

Franklin og Brozek fester seg særskilt ved at de ikke finner noen systematiske endringer i forhold til behovspersistente tendenser. «Det er overraskende,» skriver de, «at forsøkspersonenes sterke behovspersistente atferd (under matdeprivasjon) ikke kommer til syne i testskårene . . . Rosenzweigs behovspersistente svar legger vekt på løsning av den frustrerende spenning gjennom frivillig handling eller gjennom spørsmål eller krav om hjelp fra en annen person . . . En betydningsfull komponent ved våre fpp.s reaksjonsmønstre på den kroniske frustrasjon de befant seg i, var nettopp av behovspersistent karakter.» (21, s. 299).

Istedenfor å finne det overraskende at eksperimentalgruppen under underernæring ikke avvek positivt med hensyn til B-P, ligger det nærmere — ifølge Rosenzweigs opprinnelige tolkningsprinsipper og i lys av påstanden om at gruppen under underernæring var preget av lavere frustrasjonstoleranse — å spørre hvorledes avviket var med hensyn til O-D og særlig E-D.

En sammenligning med Taylors og Rosenzweigs norm-materiale viser at Franklin og Brozeks utvalg så langt fra å overgå, faktisk ligger betydelig

under disse utvalg i forhold til behovspersistente tendenser. På den annen side ligger Franklin og Brozeks utvalg betydelig over de to andre utvalg når det gjelder ekstrapunitive tendenser, mens det derimot i forbindelse med de resterende kategorier ikke er til stede noen virkelig større forskjeller.

Det manglende utslag på E-D-kategorien, både over tid og i sammenligning med andre utvalg, kan tolkes som uttrykk for at underernæring ikke har noen innvirkning på denne type reaksjoner.

En undersøkelse av Lindzey og Riechen (41), som tok sikte på å finne frem til en metode egnet til å fremkalle frustrasjon hos voksne, antyder at dette ikke er helt utenkelig. Resultatene av undersøkelsen gir visse holdpunkter når det gjelder forskjellige situasjoners frustrasjonsgrad. Mest frustrerende virker oftest hendelser som bringer et individ følelsen av å være skyldig i å hindre kamerater fra å nå et mål. Dernest kommer hendelser som gir et individ følelsen av at det ikke gjør det så godt som sine kamerater. Begge disse frustrasjonstyper gir seg utslag i en god del ego-defensiv atferd. Minst frustrerende virker hindring i å få mat og det at individet selv går glipp av å oppnå belønning. Lindzey og Riechen konkluderer bl. a. med: «De mest effektive midler til å fremkalle frustrasjon er mer de som blokkerer lærte eller sosialt betydningsfulle motiver, enn de som blokkerer organiske eller primære driv» (41, s. 23).

To måneder før eksperimentet tok til, var alle fpp. bl.a. blitt prøvd med B-F-testen. Umiddelbart etter den frustrerende situasjon ble alle retestet. «Hele gruppen med fpp. viste signifikant flere ekstrapunitive reaksjoner på begge prøver (B-F-testen og TAT) enn de hadde gjort i første testing,» skriver Lindsey og Riechen. «Endog uten å anta frustrasjons-aggresjonshypotesen kan en fra resultatene se at den eksperimentelle situasjon fremkalte signifikante endringer i normal atferd» (41, s. 22).

Ved siden av signifikante endringer i retning av større ekstrapunitive tendenser under frustrasjon fant Lindzey og Riechen også visse endringer i retning av sterkere ego-defensive tendenser, lavere GCR og lavere behovspersistens (40, s. 317). Endringene i behovspersistens viste seg å oppnå statistisk pålitelighet på 10 %-nivået.

I lys av Lindzey og Riechens resultater kan Franklin og Brozeks data tolkes under et nytt perspektiv. For det første merker en seg at hindring i å få mat sannsynligvis er en relativt mild form for frustrasjon, og at ego-defensive reaksjoner her ikke umiddelbart fremkalles. Som nevnt viste Franklin og Brozeks utvalg ved begge testinger en påfallende høyere ekstrapunitiv og en markert lavere behovspersistent tendens enn både Taylors og Rosenzweigs normalutvalg. Interessant nok er det nettopp i forbindelse med disse kategorier Lindzey og Riechen finner de største forskjeller i sitt materiale.

Det ligger ytterst nær å fremsette den hypotese at den manglende endring i testresultatene i Franklin og Brozeks undersøkelse kan tilskrives det forhold at fpp.s delvise fysiske rehabilitering ennå ikke er blitt etterfulgt av en tilsvarende psykologisk pre-eksperimentell rehabilitering.

Hvorvidt denne hypotese er holdbar, er det selvsagt umulig å gi svar på. Så lenge den ikke er blitt prøvd, synes det å være å gå noe for langt

— slik Franklin og Brozek gjør — å avskrive B-F-testen diagnostisk og klinisk verdi.

Den vesentligste svakhet ved Franklin og Brozeks undersøkelse er at den ikke tilstrekkelig rigorøst følger kravet om eksperimentell kontroll. Franklin og Brozek er selv klar over dette. De unnskylder seg med at pre-testing ikke var mulig av tekniske grunner, men de unnlater å trekke den fulle konsekvens av dette i sine konklusjoner.

Mens tidsintervallet mellom første og annen gangs testing i Franklin og Brozeks undersøkelse var ca. 3 måneder, var det hos Lindzey og Riechen 2 måneder. I begge tilfelle er tidsintervallet så stort at sannsynligheten for at fpp. husket sine tidligere svar, er svært liten.

Det samme kan ikke sies om fpp. i French's undersøkelse (22). Tidsintervallet mellom første og annen gangs testing var her bare 3 uker. Formålet med undersøkelsen var å finne frem til hvilken virkning frustrasjon og belønning har på individers reaksjonstilbøyeligheter. Som fpp. ble anvendt 80 studenter. Blant disse hadde 40 på en kurs-eksamen fått karakteren A eller B, 40 C eller D. For halvparten av de flinkeste studentene ble det opplyst at de hadde fått en karakter som lå to trinn under den de faktisk hadde fått. For halvparten av de dårligste studentene ble det på tilsvarende måte opplyst at de hadde fått en karakter som lå to trinn høyere enn den de hadde fått.

French antar at det er en nær korrelasjon mellom den karakter studentene får, og deres forventninger og aspirasjoner. Den gruppen som fikk sine karakterer redusert, antar han derfor er blitt frustrert, og den gruppen som fikk sine karakterer forbedret, er blitt belønnet.

Umiddelbart etter at karakterene ble kjent og 3 uker før eksamen, ble samtlige testet med B-F-testen.

French fremsetter ingen spesielle hypoteser han ønsker å prøve. Undersøkelsen er helt empirisk orientert med utelukkende det mål å finne frem til om testen er i stand til å differensiere mellom de forskjellige grupper.

Stort sett viser det seg å være en høy korrelasjon mellom første og annen gangs testing. Resultatene viser ingen signifikante forskjeller mellom de to grupper av flinke studenter. Derimot viser det seg at de dårlige studentene som hadde fått karakterene sine forbedret, gir signifikant flere intro-punitiv ego-defensive svar enn de som hadde fått sine riktige karakterer. Endelig viser det seg at de flinke studentene gjennomgående viser sterkere intro-punitiv behovspersistente og svakere ekstrapunitiv tendenser enn de mindre flinke studentene.

French tolker resultatene i retning av at flinke og mindre flinke studenter faktisk har forskjellige måter å møte frustrasjon på, at disse egenskaper på sett og vis hører sammen. «Det vil ikke være overraskende å finne en noe større tendens henimot intro-punitiv og behovspersistente svar blant flinke studenter,» skriver han. «En skal imidlertid være oppmerksom på at det ikke forelå noen signifikante forskjeller mellom de to grupper ved første testing: Tilsynelatende skulle det et

press slik som følger med en eksamen, til for å bringe frem disse forskjeller i reaksjonstendenser.» (22, s. 114).

French konkluderer med at alle hans resultater kan tolkes på en slik måte at de gir støtte for testens validitet. Men samtidig peker han på at det eksperimentelle og statistiske materiale er for lite til å trekke mer omfattende slutninger.

Den kanskje viktigste innvending en kan reise mot undersøkelsen, er det korte tidsintervall mellom testingene. «Hukommelse kan ha spilt inn . . .», skriver French. «Mange av svarene som ble gitt ved siste testing, er meget like de som ble gitt første gang, og mange av fpp. følte at hukommelse hadde påvirket svarene som ble gitt» (22, s. 114).

Holdbarheten av French's resultater blir i lys av dette høyst tvilsom. Faktisk reiser det tvil om selve berettigelsen av undersøkelsen.

En kritisk vurdering av de forskjellige undersøkelser som er blitt behandlet i dette kapitel, gir som resultat at vi ennå vet svært lite om B-F-testens validitet. Flere resultater peker i motsatt retning, og vi synes ennå i høy grad å famle i mørke. De aller fleste av de omtalte undersøkelser er belastet med store feilkilder, men mangelen på konsistente resultater er allikevel påfallende. Både ved sammenligning med andre tester og med manifest atferd er resultatene vanskelig å tolke. Iallfall ved første øyekast synes vi å stå overfor det forhold at B-F-testen måler *noe* forskjellig hos forskjellige personer.

Mest ekstreme med hensyn til feilkilder er Frenchs eksperimentelle undersøkelse og Du Bois og Watsons korrelasjonsanalyser. Også Franklin og Brozeks rehabiliteringsundersøkelse og Schwarz's rent empiriske analyse gir få holdepunkter. Mest opplysende er korrelasjonsberegningene med TAT — vi står her i det minste overfor forsøk på å måle det samme fenomen — og Sinaikos analyse av forholdet mellom B-F-testens forskjellige kategorier og dyktighetsvurderinger, idet vi her i det minste står overfor bestrebelse på å skaffe til veie ensartede vurderinger.

Rent prinsipielt står vi overfor to vidt forskjellige typer av tilnærmelser. TAT og andre projektive tester har som målsetting i første rekke å registrere såkalte dypereleggende personlighetslag — lag som ikke nødvendigvis faller sammen med den personen gir uttrykk for i sin daglige atferd. Sinaikos dyktighetsvurderinger hadde bakgrunn i manifest atferd, i grad av yrkesmessig suksess.

Vi vil i neste kapitel gå nærmere inn på noen av de spesielle problemer som knytter seg til disse to forskjellige tilnærmelsesformer. Vi vil fortsatt konsentrere oss om testens validitet, men belyse spørsmålet fra en mer generell synsvinkel enn den vi tidligere har nyttet.

## Kapitel IX

### Testens validitet: 4. Nivåforankring

Når en skal tolke svar på en projektiv prøve, vil det ofte være ønskelig å kjenne til hvilket atferdsnivå svarene refererer seg til<sup>11</sup>: Om de gir uttrykk for hvordan personen mener han bør handle, om hvordan han selv faktisk vil handle, eller om de gir uttrykk for mer ubevisste og latente handlingsimpulser — om personen ser på svarene sine som noe som ikke direkte har noe med ham personlig å gjøre. I første tilfelle sier vi at svarene er gitt på *opinionsnivå*, i andre tilfelle på *eksplisitt nivå*, og siste tilfelle på *implisitt nivå*.

Selvsagt vil det ofte være ønskelig å foreta en finere differensiering, kanskje særlig i forhold til det implisitte nivå — f.eks. mellom fortrenget og symbolsk, eller mellom overflatisk og dypt fortrenget nivå, men for vårt formål vil den omtalte tredeling være tilstrekkelig.

Vi vil i dette kapitel ta B-F-testens nivåforankring opp til nærmere drøfting og skal forsøke å belyse spørsmålet med så vel empiriske som teoretiske betraktninger.

#### ET HISTORISK TILBAKEBLIKK

I fremstillingen av B-F-testens historiske bakgrunn nevnte vi bl.a. at Rosenzweigs opprinnelige hypotese om et positivt samsvar mellom impulsive reaksjoner på frustrasjon og fortrenkningstendenser etter hvert ble utvidet til også å omfatte antagelsen av et positivt samsvar mellom hver enkelt av disse faktorer og hypnotiserbarhet. (Kfr. s. 12).

I en artikkel (63) gjør Rosenzweig nærmere greie for en empirisk undersøkelse av denne såkalte triadiske hypotese.

---

<sup>11</sup> Av tilsvarende betydning er det ofte å ha kjennskap til hvilken figur fpp. identifiserer seg med på bildene. I B-F-testen vil dette nesten bestandig være den talende figuren personen svarer på vegne av, men det forekommer også anormale identifiseringer, f.eks. konsistent identifisering med figurer som er i den mest fordelaktige stilling. Slike anormale identifiseringer forekommer sannsynligvis ikke særlig ofte, men hvis de forekommer og en ikke er oppmerksom på forholdet, men slavisk gjør bruk av skåringsmanualen vil skåringsprofilen naturligvis bli høyst misvisende.

Undersøkelsen er basert på et utvalg på tre grupper med henholdsvis 23, 21 og 20 studenter. Bare på den siste gruppen ble hypnose foretatt eller forsøkt. I forbindelse med to andre nøyde en seg med å registrere suggestibilitet på en måte som tidligere hadde vist seg å korrelere relativt høyt ( $r = + .52$ ) med hypnotiserbarhet. Rosenzweig presenterer samsvaret mellom de tre forskjellige variabler i form av rang-korrelasjoner, beregnet særskilt for hver enkelt gruppe.

Det empiriske materiale støtter den fremsatte hypotese. Samsvaret mellom impulsive tendenser og fortregning er konsistent positivt (.19, .39 og .54), og det samme er samsvaret mellom impulsive tendenser og suggestibilitet (.52 og .47) og hypnotiserbarhet (.78).

Undersøkelsen ble publisert i 1942. Ennå befant frustrasjonstesten seg på eksperimentalstadiet. Skåringen ble etter eksperimentallrapporten å dømme utelukkende foretatt i forhold til reaksjonsretninger, og noen presise skåringsregler eller skåringseksempler var ennå ikke utarbeidet.

Frustrasjonstesten (F-testen) besto på dette tidspunkt i alt av fire deler:

En første del (P), hvor «fp. ble konfrontert med situasjoner konstruert slik at projeksjon av ubevisste stereotype innstillinger ble favorisert.» En annen del (I), hvor fp. ble stilt overfor en serie med verbalt beskrevne frustrerende situasjoner, hver etterfulgt av tre svaralternativer, et ekstrapunitivt, et intro-punitivt og et impunitivt alternativ, — med anmodning om å angi et alternativ som mest var i overensstemmelse med hvordan hans idealperson ville handle. Testens tredje del (R) var lik annen del, bare med den forskjell at fp. nå fikk i oppdrag å angi på hvilken av de nevnte måter han selv faktisk ville reagere. Testens fjerde og siste del (B) tok utgangspunkt i mer reelt frustrerende betingelser. Fp. ble her pålagt med sterk tidsbegrensning å løse papir og blyant-prøver, bestående av labyrinter, aritmetiske og geometriske problemer. Prøvene var slik laget at svært mange av oppgavene var uløselige innenfor den tid som var tillatt. Umiddelbart etterpå ble fp. anmodet om å angi sine egne følelser — også nå ved å avmerke et av tre på forhånd oppsatte svaralternativer, svarende til ekstrapunitiv, intro-punitiv og impunitiv reaksjonstype.

Testens fire deler representerer ifølge Rosenzweig forsøk på å registrere reaksjonstendenser på forskjellige atferdsnivåer. Del P tok sikte på implisitt nivå, del I på rent opinionsnivå, del R på eksplisitt opinionsnivå og del B på et rent eksplisitt nivå.

Rosenzweigs beskrivelse av de forskjellige testdeler er svært skjematisk. Han kommer således ikke inn på hvilke verbale situasjonstyper som blir nyttet i del I og R, og heller ikke hvilke problemer og hvilken administrasjonsform som blir nyttet i del B. Heller ikke del P blir nærmere beskrevet.

Selve det metodiske opplegg er slik at det gir store muligheter for en nærmere analyse av forholdet mellom individers reaksjonstendenser på forskjellige atferdsnivåer.

Noen systematisk undersøkelse av dette spørsmål blir ikke behandlet, men det blir henvist til at en slik undersøkelse er under utvikling. Rosenzweig skriver bl. a.: «Ved å sammenligne del P med de andre delene av F-testen er det innlysende at svarene på denne del, hvor individet formodes å projisere sine implisitte stereotypier, er den mest relevante for den triadiske hypotese. Svarene i del I og R er påvirket og begrenset av instruksjonen, som pålegger fp. å angi sin oppfatning av sine reaksjoner. Del B, som ble konstruert for å måle aktuell atferd, omfatter upersonlige problemer. Studier som er i gang av Frost ved Clarke Universitet over F-testens validitet, antyder at del B i sin nåværende form er en dårlig indikator på individers reaksjoner på frustrasjon, at både del I og R måler stort sett det samme, og at reliabiliteten av disse «opinion»-svar er lav. I de fleste av sine eksperimenter har Frost funnet at del P er mest givende og også den som viser best korrelasjon med uavhengige kriterier. Av disse grunner vil vi i hovedsaken benytte oss av del P i analysen av de foreliggende resultater.»<sup>12</sup> (63, s. 9).

Rosenzweig gir ingen nærmere opplysninger om hvilke uavhengige kriterier Frost har brukt, eller hvor høye korrelasjoner han har funnet. Heller ikke gir han, som nevnt, noen opplysninger om korrelasjonen mellom de forskjellige deler av testen. Begge deler kunne ha gitt oss verdifulle holdepunkter til vurdering av hvorvidt del P faktisk registrerer implisitte reaksjonstendenser.

## NYERE UNDERSØKELSER

Mens Rosenzweig opprinnelig synes å betrakte B-F-testen som et instrument til registrering av implisitte reaksjonstendenser, har han i senere arbeider fremhevet at det som arbeidshypotese er vanlig å tillegge testen validitet på eksplisitt nivå — medmindre det foreligger særlige opplysninger som peker i annen retning (73, s. 70).

Resultatene av en undersøkelse av Lindzey og Goldwyn (42) gir støtte til Rosenzweigs endrede oppfatning. De arbeider med et utvalg på 20 studenter og tar spesielt sikte på å fastslå B-F-testens nivåforankring og korrelerer resultatene på testen med både selvtakseringer (self ratings) og takseringer av manifest og latente atferd. De konsentrerer seg utelukkende om den ekstrapunitive kategori på testen. Til taksering av manifest aggresjon bruker de detaljerte autobiografier og intervjuopplysninger, mens deres taksering av latent aggressivitet representerer en slags samlet vurdering basert på data fra forskjellige kilder — bl.a. andre testresultater, intervjuopplysninger og fantasi-produkter.

Både når det gjelder selvtaksering og taksering av manifest atferd, forsøker de å skaffe seg særskilte mål på bl.a. fysisk og verbal aggressivitet.

---

<sup>12</sup> Resultatene som tidligere er nevnt, er således basert på bruken av del P. Det er denne delen som senere er blitt betegnet Rosenzweigs billedfrustrasjonstest.



Størst korrelasjon finner de mellom testresultater og fysisk aggressivitet på det manifeste plan.

Denne korrelasjon (.42) oppnår en statistisk pålitelighet på 5%-nivået. For øvrig finner de konsistent lavere korrelasjon med verbal enn med fysisk aggressivitet, og i forbindelse med latent aggressivitet får de en mindre negativ korrelasjon (— .17). Resultatet antyder, konkluderer de, at testen «primært er sensitiv overfor manifest eller bevisst aggresjon» (42, s. 533).

Lindzey og Goldwyns konklusjon står i direkte motsetning til resultatene av en tidligere undersøkelse av Holzberg og Posner (30). Også Holzberg og Posners hovedinteresse var å finne frem til B-F-testens nivåforankring, og også de brukte forskjelligartede uavhengige atferdskriterier. Som fpp. nytter de en gruppe med 47 sykepleierskelevere. 1) Ved administrasjon av Allports *Ascendance-Submission Reaction Study* får de uttrykk for fpp.s egen vurdering av deres tendenser i retning av pågåenhet. 2) Ved bruk av atferdsobservasjoner av personer som hadde tilsyn med elevene, får de en standardisert taksering av den enkeltes grad av pågåenhet i situasjoner med tilstedeværelse av «autoriteter». 3) Gjennom en sosiometrisk tilnærming — hvor de enkelte elever ble anmodet om å vurdere de andre elevene — får de et uttrykk for den enkeltes pågåenhet i forhold til sine studiekamerater, og endelig 4) ved bruk av TAT, modifisert som gruppetest, får de et mål på den enkeltes pågåenhet på fantasiplanet. De begrenser seg for øvrig til utelukkende å beregne korrelasjonen med retningskategoriene på B-F-testen.

Ingen av deres produktmoment-korrelasjoner oppnår noen nevneverdig statistisk pålitelighet. Størst er korrelasjonen med TAT. Holzberg og Posner går enda et skritt videre i sin analyse, idet de tar for seg den høyeste og laveste femtedel av fordelingen på de enkelte retningskategorier på B-F-testen og beregner forskjellen mellom dem. De finner da at ekstrapunitive tendenser på testen diskriminerer signifikant i forhold til selvhøvdende fantasier, men ikke i forhold til noen av de andre atferdsvariablene. De konkluderer med at B-F-testen «ikke synes å gi uttrykk for aggressivitet i manifest atferd, men mer synes å gi et mål på tilstedeværelsen av aggressivitet i individets fantasi-liv» (30, s. 775).

Forskjellen mellom Lindzey og Goldwyns og Holzberg og Posners konklusjoner er påfallende. Når en skal vurdere de to undersøkelser, merker en seg kanskje først at Holzberg og Posner arbeider med et over dobbelt så stort utvalg av fpp. Et særtrekk ved Holzberg og Posners undersøkelse er for øvrig at både tilsynslærernes vurdering den sosiometriske taksering og den modifiserte TAT tar sikte på å registrere pågåenhet (assertive behavior). Svært ofte vil dette mer si positivt, utoverrettet initiativ enn direkte aggressive (ego-defensive) eller ekstrapunitive tendenser. Det virker som de atferdstrekk som blir registrert i de to undersøkelser, ikke er nøyaktig de samme. (Når

det gjelder B-F-testen, er det f.eks. nærliggende like mye å tolke den intro-punitiv-behovspersistente faktor som uttrykk for pågåenhet som noen av de ekstrapunitive enkeltfaktorene.) At dette forhold alene kan forklare forskjellene i resultater, er imidlertid tvilsomt. Hvilken av de to konklusjoner som er mest holdbar, vil selvsagt være avhengig av de forskjellige kriteriers validitet, men vi kan heller ikke se bort fra at kanskje begge konklusjoner er uholdbare, at B-F-testen ikke har noen stabil nivåforankring. Begge de omtalte undersøkelser tar mer eller mindre for gitt at testen har en fast nivåforankring, og at denne så å si kan fastslås én gang for alle. Det synes faktisk sannsynlig på bakgrunn av forskjellen mellom resultatene i de to undersøkelser at selve dette resonnement er uholdbart, at testen har en meget labil nivåforankring. I denne retning peker også de høyst divergerende resultater vi tidligere har behandlet — når det gjelder B-F-testens samsvar med kriterier på så vel latent som manifest atferd. Vi sikter i først rekke til forskjellen mellom Tejessys og Lindzey og Goldwyns TAT-korrelasjoner (kfr. s. 58), og til forskjellen mellom Simos og Lindzey og Goldwyns korrelasjoner med atferdstaksringer, i eksperimentelt tilrettelagte stress-situasjoner. (Kfr. s. 64).

## FORHOLDET MELLOM ATFERDSNIVÅER

Av generell interesse er de korrelasjonsberegninger Holzberg og Posner foretar mellom TAT-skårer og deres kriterier på manifest atferd. De finner her følgende korrelasjoner: med Allports reaksjonstest —.04, med sosiometrisk taksering —.11 og med tilsynslæreres vurdering —.31. Av disse korrelasjonene er det bare den siste som oppnår statistisk pålitelighet, men alle går i samme negative retning. De forskjellige kriterier finner de for øvrig er signifikant positivt korrelert.

«Den signifikante inverse korrelasjon mellom tilsynslæreres vurderinger og TAT,» skriver Holzberg og Posner, «støtter Symonds konklusjon at «Hvis en person arbeider ut sine problemer i åpen atferd, vil han ikke finne det nødvendig å arbeide dem ut i fantasien — og hvis han arbeider dem ut i fantasien, er han ikke bundet til å uttrykke dem i virkeligheten . . .» En kvalitativ analyse av det foreliggende materiale brakte også frem at meget få forsøkspersoner viste fravær av aggresjon både i åpen atferd og fantasi . . . Problemet er ikke tilstedeværelse eller fravær av aggresjon, men mer hvorledes individet forholder seg til den og uttrykker den.» (30, s. 773).

Ifølge Symonds skulle vi altså vente en negativ korrelasjon mellom aggresjon i fantasi og i manifest atferd. I andre undersøkelser hvor dette spørsmål her blitt tatt opp, har det fortrinnsvis vært et visst positivt eller en total mangel på korrelasjon som er blitt påpekt. I en undersøkelse av Korner blir det f.eks. ikke funnet noen korrelasjon mellom manifest og latent aggressivitet (38, 39), mens det i en undersøkelse av Sanford et al. (82) blir påvist en korrelasjon på +.15 mellom aggresjon vurdert på grunnlag av TAT og på grunnlag av manifest

atferd. I en tilsvarende undersøkelse av H. A. Murray (49) blir det derimot ikke funnet noen korrelasjon.

Både Sanfords og Murrays undersøkelser tar opp til analyse også andre atferdstrekk enn aggresjon. De finner at det er stor forskjell mellom atferdstrekk når det gjelder korrelasjon mellom atferdsnivåer. Murray påviser f.eks. at når det gjelder slike behov som passivitet, avvisning og dominering, foreligger det en positiv korrelasjon, mens det særlig i forbindelse med seksuelle behov er en negativ korrelasjon. I Sanfords undersøkelse påvises en variasjon i korrelasjon fra  $-0.44$  til  $+0.41$ .

Sanford fremsetter en sosialpsykologisk forklaring på variasjonen i korrelasjon. Han tillegger de sosiale forhold stor betydning. Er de sosiale forhold slik at de nekter bestemte behov tilfredsstillelse, vil forskjellen være stor; behovene vil gjøre seg gjeldende på fantasiplanet, men ikke i manifest atferd. Er de sosiale forhold derimot slik at de oppmuntrer bestemte behov, men samtidig ikke gir dem tilstrekkelig mulighet for tilfredsstillelse, vil behovene gjøre seg gjeldende både i fantasi og atferd.

Ut fra denne teori skulle vi vente at f. eks. forholdet mellom latent og manifest aggressivitet skulle variere fra en kulturform til en annen, og også kanskje fra en befolkningsgruppe til en annen. Både Sanford og Murray arbeider med studentutvalg, med fpp. som sannsynligvis hovedsakelig har sin forankring i middelklasse-forhold. Vi skulle vente at utvalget til Symonds hadde som bakgrunn et mer aggresjonsfiendtlig miljø enn Sanfords og Murrays utvalg, og vi skulle likeledes vente at det i et utvalg med fpp. fra arbeiderklassen, hvor aggresjon blir mer oppmuntret og mildere straffet, vil foreligge en positiv korrelasjon mellom aggressive fantasier og atferd. I en undersøkelse av Mussen og Naylor (50) blir dette sistnevnte forhold også påvist å være tilfelle.

Forholdet mellom atferdsnivåer synes med andre ord å være avhengig av generelle sosialkulturelle betingelser, men spørsmålet kan også behandles ut fra en rent personlighetspsykologisk synsvinkel.

Et standpunkt som ofte er blitt hevdet, er at personer med dårlig jeg-kontroll vil vise større samsvar mellom implisitte og eksplisitte tendenser enn andre. Symonds uttrykker det f. eks. på følgende måte:

«Det er en tankevekkende hypotese at det i emosjonelt stabile og normalt veltilpassede individer er en sterk motstand mot å uttrykke ubevisste og fortrenge impulser slik at det her er lite samsvar mellom fantasimateriale og manifest atferd — den veltilpassede person lever realistisk i det nærværende, og hans fantasier er gjenoppvekkelser av lenge undertrykte impulser og tendenser; men at konfliktene i emosjonelt forstyrrete individer presser individet til å åpenbare og gjenleve ubevisste tendenser og gi uttrykk for dem i atferd, slik at det er et nært samsvar mellom fantasimateriale og personlighet.» (93, s. 588).

Flere dynamisk orienterte psykologer har tatt sterk avstand fra de mer ortodokse synsmåter Symonds her gjør seg til talsmann for. Også Symonds selv har i sine senere arbeider i høy grad forandret stand-

punkt, idet han her hevder at jo større kontrast som foreligger, desto *dårligere* vil i alminnelighet individets tilpassing være.<sup>13</sup>

Til en viss grad avspeiler forskjellige standpunkter i dette spørsmål forskjeller i grunnsyn på psykisk sunnhet. På den ene side har vi dem som setter jegets evne til effektiv fortrenkning i første rekke, på den andre side dem som først og fremst ser det som ønskelig at det foreligger en indre integritet, en mangel på blokkering og avstengning av dype-religgende reaksjonstendenser og impulser.

At det foreligger en liten korrelasjon mellom fantasiliv og manifest atferd hos mange grupper av psykiatriske pasienter, blir støttet av empiriske undersøkelser.<sup>14</sup>

I en undersøkelse av Fisher og Hinds som tok spesielt sikte på å klargjøre forholdet mellom aggressivitet på forskjellige atferdsnivåer hos henholdsvis normale, paranoide schizofrene og schizofrene med suicidale tendenser, blir det bl.a. konkludert med at det er større samsvar mellom aggressivitet på forskjellige nivåer hos normale enn hos de to andre grupper (20, s. 29). Videre fremhever Fisher og Hinds at et misforhold mellom nivåer gir bedre differensialdiagnose med hensyn til feiltilpassing enn forskjeller på ett og samme nivå. Et vesentlig tegn på feiltilpassing synes således ofte å være en ekstrem intro-punitiv selvpoppfatning, knyttet sammen med sterke ekstrapunitiv tendenser på implisitt nivå.<sup>15</sup>

Med bakgrunn i de synspunkter som ovenfor er hevdet, skulle vi vente at en personlighetsprøve, dersom den registrerer implisitte reaksjonstendenser, skulle vise forskjellig samsvar med kriterier på manifest atferd alt etter typen av fpp. som blir anvendt. En høy korrelasjon med manifest atferd vil altså i enkelte tilfelle ikke være til hinder for at prøven måler latente trekk.

I den grad vi tar i bruk personlighetsprøver med høy nivåabilitet, blir det selvsagt vanskelig å komme frem til eksakt kunnskap om forholdet mellom atferdsnivåer. Stort sett må dybdeanalyser derfor ansees som en betydelig mer egnet forskningsmetode på dette område enn projektive prøver.

---

<sup>13</sup> «Generelt sett», skriver Symonds, «er samsvaret (mellom fantasi og karakter) best blant godt tilpassede personer. Normale, lykkelige personer fortalte realistiske historier som var fri for overdrivelser eller fordreielser... Men mindre veltilpassede personer fortalte historier som ikke svarte til deres karakter . . . I mange tilfelle ga historiene til kjenne uaksepterte tendenser ønsker eller mål; personens karakter var et panser, en beskyttelse mot å innrømme realiteten av disse tendenser. Personer som fortalte aggressive historier, var føyelige og forskremte, de med depressive historier var i virkeligheten muntre og prangende . . .» (94, s. 110).

<sup>14</sup> En ganske annen ting er at det blant pasienter av denne type ofte synes å skulle lite provosering eller frustrering til før mer infantile implisitte tendenser bryter frem og gir seg uttrykk i deres manifeste atferd.

<sup>15</sup> Det er i denne forbindelse av interesse å merke seg at Abee og Goldman blant psykiatriske pasienter fant at «i de tilfelle hvor pasientene er ekstrapunitiv og deres testskårer er sterkt intro-punitiv, viser prognosen seg å være særdeles dårlig» (kfr. s. 63).

Når vi tidligere fremhevet at B-F-testen synes å være et nivålabilt instrument, var dette med bakgrunn i en relativ betraktningssmåte. Både Holzberg og Posner, og Lindzey og Goldman arbeider med uavhengige mål på manifest og latent atferdsnivå. Hadde det ikke vært fordi TAT korrelerer negativt med manifest atferd i Holzberg og Posners undersøkelse, kunne en være fristet til å tro at forskjellen i resultater ga uttrykk for en forskjell mellom atferdsnivåer i de to utvalg. Den positive korrelasjon mellom TAT og B-F-testen, sett i forhold til den begrensede negative korrelasjon mellom B-F-testen og latent atferd i Lindzey og Goldwyns undersøkelse, antyder imidlertid sterkt at B-F-testen har en forholdsvis stor nivålabilitet.

## NIVÅBESTEMMENDE FORHOLD

Hvis vi tar som gitt at B-F-testen har en labil nivåforankring, blir det neste spørsmål hvilke forhold som her er bestemmerde. Er f. eks. testens nivåforankring forskjellig hos forskjellige typer av personer? Er administrasjonsformen av betydning?

Som tidligere nevnt (s. 56) gjorde Fisher og Hinds i sin undersøkelse også bruk av en sterkt forkortet utgave av B-F-testen. En sammenligning mellom de forskjellige utvalg når det gjelder B-F-testens korrelasjon med andre tester kan kanskje gi oss en del holdepunkter.

I sitt normalutvalg finner Fisher og Hinds størst positiv korrelasjon mellom B-F-testen og et spørreskjema (.42), i utvalget med paranoide schizofrene størst korrelasjon med åpne aggressive svar på TAT (.32), og i utvalget med suicidale, størst positiv korrelasjon med indirekte aggresjon på TAT (.26). I de to sistnevnte utvalg finner de derimot praktisk talt ikke noen korrelasjon med spørreskjemaet (.07 og .11). Resultatene til Fisher og Hinds antyder at B-F-testen har forskjellig nivåforankring hos forskjellige personlighetstyper, og at testen stort sett er mer projektiv, favner dypere personlighetslag, hos psykiatiske pasienter enn hos normale, og er mer projektiv hos suicidale enn hos paranoide schizofrene.<sup>16</sup>

Ved klinisk bruk av testen overfor psykiatrisk klientel er det ved flere anledninger blitt lagt vekt på uoverenstemmelsen mellom manifest atferd og svarene på B-F-testen. Som illustrasjon på anvendelsen av testen refererer Rosenzweig et sted til testjournalen til en paranoid pasient:

«Pasientens svar på B-F-testen var av særlig interesse fordi de synes å reflektere hans underliggende følelse av skyld og mindreverdighet mer enn hans mer symptomatiske fiendtlige atferd . . . B-F-testen brøt i dette tilfelle gjennom til noen av de dypere personlighetsmønstre, som i pasientens paranoide projeksjon kom til syne som referanse- og forfølgelsesillusjoner.» (71, s. 174).

---

<sup>16</sup> Det må bemerkes at dette er nokså løse antydninger. Fisher og Hinds's utvalg er små, korrelasjonsforskjellene er små, og deres bruk av B-F-testen innskrenker seg til 12 av leddene.

Når Rosenzweig opprinnelig hevdet at B-F-testen har en implisitt nivåforankring, men senere endret standpunkt, så kan dette skyldes at han har gjort bruk av testen overfor forskjellige typer av klientel.

Det er rimelig å anta at normale vil ha en mindre projiserende tendens, en større jeg-kontroll, som gjør det vanskeligere for dem å svare på vegne av et anonymt individ, enn mange psykiatriske pasienter. Riktignok skulle vi også vente den motsatte tendens, den nevrotiske og spontanitetshemmete personlighet, som vil være så opptatt av hva testadministratoren venter av ham at han overhodet ikke vil være i stand til å gi uttrykk for hverken sitt bevisste eller ubevisste jeg, men utelukkende sine hyperaktive kontrollmekanismer.

At ikke bare fpp.s struktur, men også selve testsituasjonen er av betydning, antyder en undersøkelse av Holzberg og Hahn (31).

Formålet med undersøkelsen var å studere hvorvidt B-F-testen er i stand til å differensiere mellom psykopater og normale. Deres psykopatutvalg består av 17 elever ved et gutteskolehjem, og deres normalutvalg av 20 elever ved en «high school». Mens alle i psykopatutvalget var preget av asosial og aggressiv manifest atferd, hadde ingen av elevene i normalutvalget noen gang vært i konflikt med politi eller skole. Elevene i de to utvalg ble ellers valt ut på en slik måte at de med hensyn til både kjønn, alder, intelligens, leseferdighet og sosialøkonomisk status var relativt lik hverandre.

Holzberg og Hahns utgangspunkt er at «Psykopatene er kjennetegnet ved aggressivitet ikke bare i atferd, men også i fantasi. Deres fantasiliv vil avsløre mer fiendtlighet enn normale.» (31, s. 776). Uansett hvilket svarnivå som anvendes, skulle en vente at psykopatene ga høyere ekstrapunitiv skårer.

Ved sammenligning av de to utvalgs resultater på B-F-testen finner Holzberg og Hahn en betydelig større spredning i psykopatutvalget, mens derimot gjennomsnittsverdiene i de to utvalg ikke avviker på noen statistisk pålitelig måte. De tolker dette som uttrykk for at B-F-testen har en høyst begrenset diskrimineringssevne.

Vi kan selvsagt ikke direkte bestride holdbarheten av en slik tolkning. Det ligger imidlertid meget nær å stille spørsmålet om hva grunnen kan være til at testen ikke diskriminerer i et tilfelle som dette.

Deres undersøkelse stopper ikke med en systematisk sammenligning av de to utvalg. De forsøker seg også på sammenligninger med utgangspunkt i tre forskjellige modifiseringer av testen. Deres resultater i denne forbindelse åpner interessante perspektiver. Etter først å ha nyttet testen i sin standard-form administrerer de den senere, én gang med endret tekst, slik at graden av frustrasjon i første halvpart av situasjonene blir økt, — én gang med endrede bilder, idet ytterligere figurer med oppmerksomheten henvendt på den frustrerte person blir introdusert i halvparten av situasjonene, — og endelig én gang med endret instruksjon, med oppfordring om å svare hva personen på bildene ikke vil si, men faktisk ville like å si. Fpp. som ble nyttet, var hele tiden de samme, og tidsintervallet mellom hver testing 3 uker.

Den førstnevnte endring tok sikte på å undersøke hvorvidt testen er sensibel overfor skyldfølelse i forbindelse med aggresjon. Ved å øke frustrasjonen i første halvpart av testen antok de at antallet av aggressive svar ville bli økt. Og ved å sammenligne svarene i siste halvdel med svarene på standardformen antok de videre at de ville få et uttrykk for skyldfølelse: Jo større reduksjon i aggressive svar i forhold til standardformen, desto mer skyldfølelse.

I overensstemmelse med sine antagelser finner de at psykopatutvalget viser signifikant mindre tendens til å redusere sin aggressivitet, og dermed at det sannsynligvis er mindre preget av skyldfølelse.

Hensikten med deres annen endring — økningen av antall tilstedeværende figurer på halvparten av bildene — var å undersøke hvorvidt de to utvalg ville reagere forskjellig, stilt overfor «symboler for sosial autoritet». Psykopater, antok de, vil vanligvis i slike situasjoner i større grad enn andre redusere sin aggressivitet og så å si «spille» ikke-aggressive.<sup>17</sup>

Deres resultater svekker en slik hypotese. Psykopatutvalget viser ingen signifikant endring (i forhold til standardformen). Derimot finner de en signifikant aggresjonsøkning i normalutvalget.

Deres tredje endring av testen er basert på ønske om å få mer fantasipregede svar enn det er rimelig å få på grunnlag av standardinstruksjonen. Med bakgrunn i sine teoretiske overveielser vedrørende psykopati antar Holzberg og Hahn at de i dette tilfelle skulle få høyere ekstrapunitiv skårer i psykopatutvalget enn i normalutvalget.<sup>18</sup>

Heller ikke denne hypotese blir støttet. På fantasiplanet — hvis vi antar at det er dette plan som i siste tilfelle fortrinnsvis blir registrert — viser det seg at normalutvalget faktisk skårer signifikant mer ekstrapunitivt enn psykopatutvalget.

Istedenfor å forsøke å finne en naturlig forklaring på forholdet foretrekker Holzberg og Hahn å se på resultatet som enda en antydning om B-F-testens manglende validitet.

I høy grad må resultatet sies å støtte Symonds hypotese — at det er et negativt samsvar mellom aggressivitet i fantasiliv og manifest atferd. Psykopatutvalgets mer aggressive manifeste atferd kan vanskelig betviles. Det uløste spørsmål er imidlertid hvorfor dette ikke ga seg markerte utslag ved sammenligningen av de to utvalgs resultater på standardformen av testen.

Det er mulig Holzberg og Hahn er inne på en forklaring når de innledningsvis skriver: «Psykopatene vil hemme aggressiv atferd når de står overfor ytre symboler for sosial autoritet. Under betingelser hvor deres atferd

<sup>17</sup> Deres antagelse her er høyst tvilsum. En vanlig oppfatning av psykopater går ut på at de er svært ufølsomme for konsekvensene av sine handlinger, og at de bare under ytterst konkret sosialt press er i stand til å modifisere sin atferd og midlertidig hemme sin frittflytende aggressivitet. En økning av antall tilstedeværende personer vil neppe representere et tilstrekkelig konkret sosialt press.

<sup>18</sup> Også denne antagelse kan diskuteres. Det blir f.eks. ofte antatt at psykopater har mer angst overfor indre aggressive fantasier enn ytre konsekvenser av aggressive handlinger. Aggressive fantasier blir derfor øyeblikkelig overført i handling, og fantasilivet forblir tilsynelatende uberørt.

blir gjort til gjenstand for kritisk vurdering av andre, vil de redusere sin aggressivitet til et punkt som gjør deres atferd tilsynelatende identisk med normales. . . . Psykopatene er vel kjent med samfunnets krav, men de har ikke innforlivet (internalized) dets sosiale verdier i deres personlighetsstruktur.» (31, s. 776).

Med dette som bakgrunn er det klart at Holzberg og Hahn ikke har foretatt noen helgardering. Selve testsituasjonen representerer en betingelse hvor «atferd blir gjort til gjenstand for kritisk vurdering av andre». I den grad psykopatutvalget har vært influert av testsituasjonen og gitt sine svar på et opinionsorientert nivå, skulle vi faktisk ikke vente å finne forskjeller mellom utvalgene.

Det er et par forhold som kan tyde på at dette er tilfelle. Ved å øke antall tilstedeværende figurer på halvparten av bildene håpet Holzberg og Hahn å innføre «symboler på sosial autoritet». Mangelen på reduksjon av aggressivitet i psykopatutvalget kan tolkes å antyde at psykopatutvalget er opinions- og konformitetsorientert ved begge betingelser, mens aggresjonsøkningen i normalutvalget kan antyde at dette utvalg først ved siste betingelse ble influert av ideelle overveielser, av hva de antok ble ventet av dem av deres omgivelser — og da kanskje i første rekke av deres kamerater.

Et annet forhold som kan støtte en slik forklaring, er den uttalte forskjell som sannsynligvis er til stede i de to utvalgs testsituasjon. Psykopatutvalget befant seg i en skolehjemssituasjon, en situasjon som trolig i høy grad har virket til å gjøre mange av fpp. svært mistenksomme og kritiske overfor testen. Den store spredning i svarfrekvensen på de enkelte kategorier kan bl.a. tolkes i denne retning.

En vet selvsagt ikke i hvor høy grad de synspunkter som ovenfor er blitt fremsatt, dekker de faktiske forhold.<sup>19</sup> At de sterkt antyder at testsituasjonen er av betydning, kan neppe betviles.<sup>20</sup>

Et annet spørsmål som dermed melder seg, er om testsituasjonen er av den samme betydning for forskjellige personlighetstyper. Det er høyst rimelig å anta at dette ikke er tilfelle. Vi står i så fall overfor det forhold at svarnivå på B-F-testen er avhengig av et vekselspill mellom ytre og indre betingelser, av interaksjonen mellom personlighetstype og testsituasjon. Det empiriske materiale som i dag foreligger om testen, synes å støtte holdbarheten av et slikt standpunkt.

---

<sup>19</sup> Det forhold at de samme forsøkspersoner ble anvendt hver gang og dermed fikk testen hele 4 ganger i løpet av 3 måneder, kan i høy grad ha medvirket til en kunstig retest-konsistens, og til redusert virkning av de eksperimentelle manipuleringer. En kan heller ikke se helt bort fra den mulighet et hukommelse av tidligere prøvinger kan ha gjort seg gjeldende på forskjellig måte i de to utvalg.

<sup>20</sup> Vi har også tidligere vært inne på testsituasjonens mulige betydning. I en undersøkelse med barneutgaven av testen er det blitt påvist et positivt samsvar mellom testskårer og tilpassing til en skolehjeminstitusjons krav (kfr. s. 42).



## FORELØPIG KONKLUSJON

B-F-testen er slik konstruert at den til en viss grad appellerer til forskjellige svarnivåer.

Fp. blir instruert om å svare det aller første som faller ham inn. Bildene er med hensikt tegnet slik at alle individuelle trekk ved de personer som opptrer, er utelatt. Dette er faktorer som sannsynligvis vil ha en tendens til å fremkalle latente innstillinger og følelser. Fp. vil ikke svare på egne vegne, men på det anonyme individs.

Situasjonene er imidlertid dagligdagse og ofte forekommende. Dette vil trolig lett fremme kritikk og realitetsperspektiv, slik at fp. svarer ut fra sin vanlige måte å reagere på.

Det at situasjonene eller testbildene blir presentert i fast rekkefølge, med instruksjon etc., innfører enda en faktor. Fp. vil lett få følelsen av å bli prøvd. Svarene vil dermed lett bli ekstra kritiske, og bli sensurert i overensstemmelse med hva som ansees riktig og sosialt akseptabelt.

Enkelte personer er mer på vakt i så henseende enn andre. Likeledes er det enkelte situasjoner som mer enn andre fremmer en slik på-vakt-holdning. Det er rimelig å anta at det er tilfelle f. eks. ved en skole-hjemssituasjon, en fengselssituasjon og enhver situasjon hvor fpp. har følelsen av at svar av en bestemt type vil bringe spesielle fordeler eller ulemper.

Uten kjennskap til hvilket svarnivå som er blitt nyttet, er det vår oppfatning at B-F-testen har en meget begrenset anvendbarhet. Brukt alene i psykodiagnostisk arbeid vil den sannsynligvis kunne gi høyst misvisende og uholdbare resultater.

Testens verdi synes i høy grad å være avhengig av at det blir utviklet spesielle metoder til registrering av svarnivå i det enkelte tilfelle. Før slike metoder er utprøvd, er det etter vår mening uforsvarlig å ta testen i bruk i rutinemessig diagnostisk arbeid.

Også Rosenzweig selv synes å være oppmerksom på at testen kan ha forskjellig nivåforankring. Etter å ha fremsatt sin arbeidshypotese om testens eksplisitte forankring skriver han: «Samtidig som en må være klar over at en ikke har noen garanti for at denne arbeidshypotese er riktig, blir det derfor pålagt den omhyggelige prøveleder å søke fra alle tilgjengelige kilder supplerende opplysninger om hvilket svarnivå fp. nytter seg av i besvarelsen av prøven» (73, s. 70).

Noe systematisk arbeid med sikte på å finne frem til en hendig metode til nivåbestemmelse er ennå ikke blitt utført. I betraktning av hvilken avgjørende betydning dette synes å ha for testens anvendbarhet, åpner det seg her høyst sentrale forskningsoppgaver.

Det som i første rekke gjør B-F-testen til et ytterst løfterikt instrument, er dens konsentrasjon om en side ved personligheten som ofte blir lite direkte berørt av andre projektive prøver, samtidig som dens administrasjonstid er kort og dens differensieringsevne er relativt god. B-F-testen synes derfor å ha tilstrekkelige muligheter til å forsvare den tid og det arbeid som vil være nødvendig for å bringe større klarhet over dens validitet.

## Kapitel X

# Egne empiriske bidrag

Vi skal i dette kapitel gjøre rede for endel empiriske undersøkelser over den norske revisjon av B-F-testen. Som nevnt ble denne testen opprinnelig konstruert som et eksperimentalsykologisk instrument. Senere er den også blitt tatt i bruk i praktisk psykodiagnostisk arbeid. Det foreligger i dag en hel rekke kasuistiske beskrivelser hvor testen er blitt anvendt (5, 7, 29, 37, 66, 67, 71, 88). I de aller siste år har den også vakt stor interesse blant sosialpsykologer.

Et sosialpsykologisk problemområde som har tiltrukket seg stor oppmerksomhet, er forholdet mellom majoritets- og minoritetsgrupper. At B-F-testen kan komme til anvendelse på dette område, antyder den tidligere omtalte undersøkelse av McCary (kfr. s. 43 f.). I en undersøkelse av Brown (8) blir B-F-testen omarbeidet spesielt med sikte på å studere sosiale innstillinger, idet enkelte av personene på testbildene blir tillagt kjennetegn slik at de oppfattes som representanter for minoritetsgrupper. Browns modifiseringer er ganske nylig blitt gjort til gjenstand for ytterligere modifiseringer av Sommer (92). Både bilder og tekst er blitt gjort mer flertydige samtidig som testbildenes antall er blitt redusert.

Et spørsmål som lenge har opptatt sosialpsykologer, er muligheten av å anvende enkle projektive metoder i omfattende rundspøringsundersøkelser. I en undersøkelse av Sanford og Rosenstock (81) blir det anvendt en del av bildene i B-F-testen nettopp for å belyse dette spørsmål. Sanford og Rosenstock konkluderer med at bruk av billedstimuli i feltintervjuing ofte synes å gi mer psykologisk meningsfulle svar enn bruk av direkte spørsmål. Likeledes hevder de at deres resultater antyder at svar materialet ved bruk av billedstimuli kan behandles med en tilfredsstillende grad av objektivitet, og at også retest-reliabiliteten ligger meget godt an sammenlignet med konvensjonell intervju-teknikk. «Det faktum at projektive data lett kan bli innhentet på dørterskelen,» skriver de, «betyr at vanlige rundspøringsundersøkelser nå kan frigjøre seg fra den begrensning som ligger i fremgangsmåter som utelukkende tar hånd om særdeles bevisste og sensurerte svar» (81, s. 15).

Det er meget som taler for at anvendelsen av projektive prøver i sosialpsykologisk forskning vil øke sterkt i årene som kommer (kfr. 1, ss. 462-485). Vårt eget utgangspunkt da vi opprinnelig tok B-F-testen i bruk, var således i høy grad rent sosialpsykologiske problemstillinger.

## BAKGRUNNEN FOR EGNE UNDERSØKELSER

Et forskningsområde som i dag står i sentrum for mange sosialpsykologers interesse, er forholdet mellom personlighet og rasehat. Det er av flere blitt hevdet at det her foreligger en høy korrelasjon, mens andre har stilt seg mer tvilende, og har som forklaring på rasehat i første rekke henvist til sosiologiske forhold, klassesetilhøring, samspillet mellom grupper, nasjonal identifisering etc.

Et problemområde som på mange måter er nær beslektet, er forholdet mellom personlighet og internasjonale holdninger. Også på dette området er det stor uenighet mellom de som støtter seg til overveiende sosiologiske forklaringsprinsipper, og de som er tilbøyelig til å hevde mer personlighetsorienterte standpunkter.

Våre egne undersøkelser med B-F-testen hadde som utgangspunkt å studere forholdet mellom personers stillingtagen i storpolitikken og deres reaksjonsmåter i personlige forhold.

Ett av de spørsmål som opprinnelig ble stilt, var å undersøke korrelasjonen mellom individers reaksjonstendenser på dagligdagse frustrasjoner og deres reaksjonspreferanser med referanse til deres nasjon i internasjonalt frustrerende situasjoner. Til registrering av de forskjellige variabler ble utarbeidet særlige innstillingsskalaer.

Som innstillingsskala til registrering av interpersonlige reaksjonstendenser, ble nyttet en serie med 40 hypotetiske, verbalt skisserte frustrerende situasjoner av dagligdags karakter. Etter hver situasjon var angitt 6 forskjellige reaksjonsmåter. Disse var valt ut og preskåret etter Rosenzweigs klassifiseringsskjema, men med den forskjell at hindringsdominans var utelatt. Fpp. ble anmodet om blant reaksjonsalternativene til hver situasjon å angi på hvilken av de nevnte måter de mest sannsynlig og dernest mest sannsynlig selv ville reagere. Ved å gi den mest sannsynlige reaksjonsmåte vekt-tallet to og den nest mest sannsynlige vekt-tallet én, håpet vi å ta hensyn til fpp.s eventuelle hierarkiske organisasjon av innstillinger eller reaksjonsdisposisjoner — og dermed oppnå relativt stabile mål på de egenskaper vi ønsket å registrere.

Til registrering av internasjonale reaksjonstendenser ble det utarbeidet en tilsvarende skala, idet vi i denne forbindelse konstruerte 40 hypotetiske internasjonale situasjoner som alle impliserte en eller annen form for frustrasjon av norske nasjonale interesser. Fpp. ble her ikke anmodet om å angi hvorledes de selv ville reagere, men hvorledes de foretrakk at deres nasjon skulle reagere i den foreliggende situasjon.

I tillegg til den interpersonlige innstillingsskalaen tok vi også i bruk Rosenzweigs B-F-test. Instruksjonen og billedteksten forelå fra før oversatt til norsk, mens derimot bildene ikke hadde fått noen norsk

utforming. Vi tok derfor initiativet til å få disse modifisert i overensstemmelse med norske forhold.

Grunnen til å trekke inn B-F-testen i undersøkelsen var håpet om at den ville gi data om reaksjonsdisposisjoner på et dypereliggende, latent atferdsnivå. Vi håpet derved å få belyst nærmere hvorvidt internasjonale innstillinger representerer projeksjoner av dypereliggende atferdstrekk, hvorvidt fortrenkte aggressive innstillinger f.eks. har en tendens til å gå sammen med aggressive internasjonale holdninger. Likeledes håpet vi å få materiale til belysning av innstillingskalaens validitet.

Det ligger utenfor rammen av dette arbeid å ta nærmere opp til drøfting de resultater som ble funnet med hensyn til forholdet mellom interpersonlige og internasjonale innstillinger. Dette materiale vil bli rapportert senere i form av en spesiell publikasjon.

Vi vil her begrense fremstillingen til å omfatte en diskusjon av B-F-testen. Vi vil utelukkende behandle den del av vårt materiale som er direkte relevant for en nærmere forståelse og vurdering av denne prøve. Vi vil derfor ikke presentere noen sosialpsykologisk studie, men konsentrere oss om en behandling av testens skåringsreliabilitet, dens indre konsistens, forholdet mellom forskjellige kategorinormer, forholdet mellom testresultater og atferdsvurderinger, og endelig til slutt forholdet mellom testresultater og svarene på vår interpersonlige innstillingsskala. Et fremtredende trekk ved store deler av fremstillingen vil være en videre diskusjon av testens nivåforankring, knyttet sammen med forsøk på en empirisk analyse av dette spørsmål.

Datainnsamlingen fant sted i august—september 1952. Som fpp. ble nyttet 167 unge menn. Av disse var 103 aspiranter ved Krigsskolen og 64 kadett- eller offisersaspiranter ved Sjøkrigsskolen. Over 80 % av utvalget var mellom 21 og 24 år, gjennomsnittet ca. 23 år. Et felles trekk ved utvalget var med andre ord interesse for en militær utdanning og en viss usikkerhet om hvorvidt utdannelsesplanene ville la seg realisere.

Undersøkelsen ble ikke knyttet sammen med noe uttaksprogram. Aspirantene ved Sjøkrigsskolen arbeidet under anonyme forhold, aspirantene ved Krigsskolen under forsikring om at alle svar ville bli behandlet strengt konfidensielt og at de ikke ville ha noen betydning for opptagelse ved skolen.

B-F-testene ble administrert som gruppeprøve — som en del av et større prøvebatteri. Prøvetiden for testen varierte en del, gjennomsnittlig lå den på omkring 25 min. Også gruppe-enhetene varierte noe. Gjennomsnittlig var det ca. 15 personer i hver gruppe. Hver gruppe gjennomgikk det samme prøvebatteri, og administrasjonsformen var hele tiden den samme.

## UNDERSØKELSE AV SKÅRINGSRELIABILITETEN

Testen viste seg å være lett å administrere. En del vansker oppsto imidlertid i forbindelse med skåringen. Et problem som straks reiste

seg, var hvorvidt vi skulle legge skåringseksemplene i den reviderte amerikanske skåringsmanual til grunn for skåringen, eller om vi skulle frigjøre oss helt fra denne og utelukkende legge vekt på de eksplisitte skåringskriteriene.

Vi bestemte oss for å ta utgangspunkt i spørsmålet om hvorvidt en direkte oversettelse av de amerikanske skåringseksemplene ville kunne tjene norske forhold. Vi gikk frem på følgende måte: Først lot vi testmaterialet bli skåret av to uavhengige personer, hvor den ene fikk instruksjon om nøye å følge skåringseksemplene, den andre om mest mulig å se bort fra disse og legge vekt på de prinsipielle skåringskategorier. Dernest ble alle svar som det ikke var oppnådd enighet om, skåret på uavhengig grunnlag av en tredje person — uten bruk av skåringseksempler.<sup>21</sup>

Hvis denne tredjeperson fortrinnsvis skulle vise seg å skåre i overensstemmelse med personen som fulgte skåringseksemplene — og totalenigheten mellom 2 av 3 var relativt høyt, ville vi anse det som forsvarlig å bruke en direkte oversettelse av skåringseksemplene. Viste derimot den tredje person fortrinnsvis enighet med personen som ikke nyttet skåringseksemplene, ville bruken av disse under norske forhold være mer betenkelig.

I de tilfelle den tredje person ikke skåret i overensstemmelse med noen av de to andre, ble det besluttet å ta disse ledd opp til samlet drøfting og endelig skåring. Ellers ble den skåre beholdt hvor 2 av 3 var enige.

Når det gjelder selve skåringen, fant denne sted etter Rosenzweigs prinsipper, bortsett fra at vi hele tiden la vekt på å unngå oppdelte skårer.

Tabel 6 gir en oversikt over den prosentvise enighet mellom de 2 skårere og antall prosent skåret likt av 2 av 3 skårere.

Av tabellen fremgår at den gjennomsnittlige enighet mellom skåring med hovedvekt på henholdsvis skåringseksempler og kriterier, er 72 %. Overensstemmelsen med reliabilitetsundersøkelsen, prosjekt 1, til Clarke, Rosenzweig og Flemming er påfallende (14, s. 366). Den gjennomsnittlige enighet mellom 2 av 4 skårere var her 74 %, og enigheten mellom 2 av 3, 90 % (kfr. s. 27). I vår undersøkelse er enigheten mellom 2 av 3, 92 %. I det store og hele må dette sies å gi uttrykk for relativt tilfredsstillende skåringsreliabilitet, fullt på høyde med resultater fra tilsvarende undersøkelser over f.eks. TAT (28).

En nærmere gransking av de forskjellige typer av uenighet som er til stede mellom de to skårere, viser at mesteparten av uenigheten i skåringen kan tilskrives forskjeller utelukkende i *enten* retning *eller*

---

<sup>21</sup> Som skårere ble anvendt to viderekommende psykologistudenter. De hadde begge gjennomgått test- og eksperimentalkurs. Ingen av dem hadde noen forutgående praktisk erfaring i bruk av B-F-testen. Deres bakgrunn og synspunkter var i første rekke preget av en akademisk psykologisk trening. Før skåringen tok til, ble de begge gitt en grundig orientering i B-F-testens konstruksjon, skåringsprinsipper og teoretiske grunnlag. Som tredje person fungerte forfatteren.

TABELL 6

Oversikt over skåringsreliabiliteten ved forskjellige testledd

| Ledd nr. | Pst. skåret<br>likt av 2 | Pst. skåret<br>likt av 2 av 3 | Ledd nr. | Pst. skåret<br>likt av 2 | Pst. skåret<br>likt av 2 av 3 |
|----------|--------------------------|-------------------------------|----------|--------------------------|-------------------------------|
| 1        | 88                       | 99                            | 13       | 57                       | 94                            |
| 2        | 68                       | 88                            | 14       | 78                       | 96                            |
| 3        | 74                       | 93                            | 15       | 70                       | 95                            |
| 4        | 86                       | 99                            | 16       | 86                       | 97                            |
| 5        | 75                       | 90                            | 17       | 44                       | 80                            |
| 6        | 70                       | 91                            | 18       | 86                       | 96                            |
| 7        | 83                       | 96                            | 19       | 78                       | 84                            |
| 8        | 41                       | 85                            | 20       | 76                       | 94                            |
| 9        | 72                       | 94                            | 21       | 78                       | 93                            |
| 10       | 78                       | 88                            | 22       | 65                       | 92                            |
| 11       | 78                       | 95                            | 23       | 53                       | 85                            |
| 12       | 58                       | 94                            | 24       | 77                       | 95                            |
|          |                          |                               | M        | 72                       | 92                            |

form. Størst uenighet foreligger i forbindelse med skåringen av reaksjonsformer (11 %). Dernest følger uenighet i forbindelse med skåring av reaksjonsretninger (8,9 %) og i vurdering av skårbarhet (4,7 %). Minst uenighet skyldes forskjellig skåring av både form og retning (3,9 %).

I alt ble 1091 av et totalantall på 3888 svar skåret forskjellig. Disse ble så senere skåret uavhengig av en tredje person. I 28 % av tilfellene viste denne skåringen uoverensstemmelse med begge de to første, i 36,7 % med ham som så vidt mulig la skåringseksemplene til grunn for skåringen, og i 35,3 % med ham som i første rekke la vekt på skåringskriteriene.

Enigheten med hver av de to første skårere er omtrent den samme. Dette antyder at bruken av de amerikanske skåringseksemplene i norsk oversettelse kan være av verdi. Hadde den tredje skårer også holdt seg til skåringseksemplene, ville sannsynligvis overensstemmelsen i dette tilfelle vært høyere, men dette ville ikke ha gitt oss noen kontroll over hva vi kan kalle skåringens «riktighet». Viktigere enn høyest mulig overensstemmelse i skåringen er at den finner sted i overensstemmelse med prinsipielle skåringskriterier.

Under ellers like forhold er det en fordel å bruke skåringseksempler, fordi dette i høy grad fører til en vesentlig innsparing av tid.

Hvis vi tar enighet mellom to av tre som uttrykk for «riktighet», følger det at skåring med hovedvekt på kriteriene ville gitt 81,8 % riktig skåring, bruken av eksemplene 82,3 %.

Av en viss interesse er det å legge merke til at det stort sett er ved de ledd hvor det er størst uenighet mellom de to første skårere, det også er størst uenighet mellom samtlige 3 skårere ( $\rho = .68$ ). Likeledes at det er en ganske god korrelasjon mellom «ledd-reliabiliteten» i vår og i Clarke, Rosenzweig og Flemmings undersøkelse ( $\rho = .65$ ).

Enkelte ledd synes altså generelt å gi relativt flertydige svar. Særlig gjelder dette ledd nr. 8 og 17. En utskifting av disse ledd til fordel for andre, mer skåringsreliable ledd vil gi testen som helhet en til dels betydelig høyere skåringsreliabilitet. Denne fremgangsmåte vil være langt å foretrekke fremfor å utarbeide definatoriske skåringseksempler. Ikke bare vil vi dermed få en høyere skåringsreliabilitet, men sannsynligvis også en høyere indre reliabilitet eller konsistens.

## UNDERSØKELSE AV INDRE KONSISTENS

En tests reliabilitet vil alltid til en viss grad være avhengig av hvilket utvalg fpp. vi arbeider med. Dette vil være tilfelle når det gjelder skåringsreliabiliteten og også når det gjelder reliabiliteten i betydningen indre konsistens.

Som tidligere nevnt påviste Taylor at 5 av testens 6 skåringskategorier hadde evne til å differensiere mellom individer på en signifikant måte (kfr. s. 33). De enkelte kategoriens indre konsistens var imidlertid ikke overbevisende høy, skåreverdienes nøyaktighet var relativt lav, og testens verdi for individuell vurdering ble sterkt betvilt.

Taylors utvalg omfattet 260 amerikanske menn. Utvalget er ganske stort, men gir allikevel ingen sikre holdepunkter for en vurdering av testens skikkethet under norske forhold. For å få dette nærmere klarlagt ble det besluttet å foreta en parallell norsk reliabilitetsundersøkelse av testen, idet vi også bestemte oss for å gjøre bruk av Hoyts variansanalytiske metode (35, s. 134).

Til reliabilitetsberegningen ble nyttet alle protokoller hvor samtlige testledd var blitt besvart og hvor alle svar var skårbare ( $N = 90$ ), og 46 av protokollene med ett manglende eller uskårbart svar. Alle protokoller med to eller flere manglende eller uskårbare svar ble utelatt. I protokollene med ett manglende eller uskårbart svar ble tilpassede skåreverdier innsatt ved bruk av Yates modifikasjon av Allan og Wisharts beregningsformel for mistede svar (91, s. 268). Tilsammen fikk vi derved et utvalg på 136 'komplette' journaler.

Tabell 7 og 8 gir en oversikt over to trinn i vår variansanalyse, og tabell 9, en oversikt over resultatene av analysen.

Tabell 9 viser at de forskjellige kategoriens indre konsistens varierer en god del. Lavest er In-kategorien med  $r_{tt} = .13$ , høyest Ex-kategorien med  $r_{tt} = .53$ . Den gjennomsnittlige reliabilitet er .34. Dette er pussig

TABELL 7

Oversikt over kvadratsummene

| Kilde    | df.*    | Ex      | In      | Im      | H-D     | E-D     | B-P     |
|----------|---------|---------|---------|---------|---------|---------|---------|
| Total .. | 3263—46 | 814.784 | 600.851 | 652.286 | 470.460 | 806.402 | 645.041 |
| Ledd ..  | 23      | 255.100 | 263.970 | 246.206 | 118.622 | 239.424 | 197.018 |
| Personer | 135     | 47.992  | 16.352  | 27.579  | 20.210  | 36.827  | 32.416  |
| Rest ..  | 3105—46 | 511.692 | 320.529 | 378.501 | 331.628 | 530.151 | 415.607 |

\* df. — degree of freedom (frihetsgrad).

TABELL 8

Oversikt over variansene (kvadratsum/df.)

| Kilde         | Ex      | In      | Im      | H-D    | E-D     | B-P    |
|---------------|---------|---------|---------|--------|---------|--------|
| Total . . . . | .2533   | .1868   | .2028   | .1462  | .2507   | .2005  |
| Ledd . . . .  | 11.0913 | 11.4769 | 10.7046 | 5.1575 | 10.4097 | 8.5660 |
| Personer ..   | .3555   | .1211   | .2043   | .1497  | .2728   | .2401  |
| Rest . . . .  | .1673   | .1048   | .1237   | .1084  | .1733   | .1359  |

TABELL 9

Oversikt over de forskjellige kategoriens reliabilitet ( $r_{tt}$ )  
og standardfeil til målingen ( $\sigma_e$ )\*

|                      | Ex   | In   | Im   | H-D  | E-D  | B-P  |
|----------------------|------|------|------|------|------|------|
| $r_{tt}$ . . . . .   | .529 | .135 | .395 | .276 | .265 | .434 |
| $\sigma_e$ . . . . . | 1.95 | 1.54 | 1.67 | 1.57 | 1.98 | 1.75 |

\* Standardfeil til målingen ( $\sigma_e$ ) er beregnet ved formelen  $\sigma_e = \sigma \sqrt{1 - r_{tt}}$ . Standardavviket ( $\sigma$ ) til fordelingen ved de forskjellige kategorier er angitt i tabell 11.



nok nesten den samme gjennomsnittsverdi som Taylor fant for sitt samlede utvalg (kfr. s. 33).

Også når det gjelder de enkelte kategorier, er det en ganske stor overensstemmelse til stede. Størst er forskjellen i forbindelse med In- og H-D-kategoriene. Mens Taylor fant en reliabilitetskoeffisient for H-D på .10, er vår .28. Og mens vi har fått en reliabilitetskoeffisient for In på .13, fant Taylor her en  $r_{tt}$  på .22.

De enkelte kategoriens evne til å differensiere mellom personer kan som tidligere nevnt undersøkes ved å signifikansprøve forholdet mellom personvariansen og feilvariansen. Tabell 10 gir en oversikt over de forskjellige kategoriens differensieringsevne.

TABELL 10

*Statistisk signifikans (p) av forholdet (F) mellom person- og restvariansen ved de forskjellige kategorier\**

| Retningskategori      | F    | p   | Formkategori        | F    | p   |
|-----------------------|------|-----|---------------------|------|-----|
| Ekstrapunitiv . . . . | 2.12 | .01 | Hindringsdominans   | 1.38 | .01 |
| Intropunitiv . . . .  | 1.16 | —   | Egodefens . . . . . | 1.57 | .01 |
| Impunitiv . . . . .   | 1.65 | .01 | Behovspersistens .. | 1.77 | .01 |

\* For  $\infty$  df.:  $F_{.05} = 1.24$   
 $F_{.01} = 1.36$

Tabellen viser at unntatt for den intro-punitive kategori, er  $F$ -forholdet for alle kategorier signifikant på .01-nivået. Sikrest differensierende er den ekstrapunitive kategori. Dette er i overensstemmelse med Taylors resultater. Også han fant Ek-kategorien å være signifikant på .01-nivået. Derimot fant han at de andre kategorier (unntatt H-D) bare nådde .05-nivået. Våre resultater er derfor mer oppmuntrende enn Taylors, på én unntagelse nær. Vår In-kategori når ikke .05-nivået. Denne kategori er ifølge våre data ute av stand til å differensiere mellom personer på en statistisk tilfredsstillende måte.

Når det gjelder graden av tillit som kan tillegges våre registrerte skårer, kan vi få et mål på dette ved hjelp av standardfeilen til målingen (kfr. tabell 9). Vi kan anta en normalfordeling av oppnådde skårer rundt en sann skåre, og spørre hvor mange prosent sannsynligvis vil ligge i øvre og nedre kvartil, hvis den sanne skåre ligger ved skårefordelingens aritmetiske gjennomsnitt<sup>22</sup>. Eller vi kan f. eks. spørre

<sup>22</sup> Hvis skårefordelingen er normal, vil  $M \pm 0,675 \sigma$  omfatte 50 % av fordelingen. Spørsmålet blir derfor hvor hyppig vil en sann skåre ved  $M$  svare til registrerte skårer utenfor  $M \pm 0,675\sigma$ . Standardfeilen til målingen ( $\sigma_e$ ) gir et hypotetisk uttrykk for spredningen av de registrerte skårer i forhold til en sann skåreverd. Jo større  $\sigma_e$  er i forhold til  $0,675\sigma$ , desto større er den del av de registrerte skårer som vil ligge i øvre og nedre kvartil hvis den sanne skåre ligger ved  $M$ .

hvor mange prosent sannsynligvis vil falle i øvre kvartil hvis den sanne skåre ligger på grensen til nedre.

TABELL 11

*Oversikt over graden av tillit som kan tillegges registrerte skårer*

| Kategori                  | $\sigma_e$ | $\sigma$ | $0.675 \sigma$ | Sann skåre ved $M$ , ant. pst. i øvre og nedre kvartil | Sann skåre ved grensen til øvre kvartil, ant. pst. i nedre |
|---------------------------|------------|----------|----------------|--|--|
| Ekstrapunitiv ..          | 1.95       | 2.91     | 1.96           | 32 %   | 2 %  |
| Intropunitiv ..           | 1.54       | 1.70     | 1.15           | 44 %   | 7 %  |
| Impunitiv . . . .         | 1.67       | 2.21     | 1.49           | 36 %   | 3 %  |
| Hindrings-<br>dominans .. | 1.57       | 1.89     | 1.27           | 42 %   | 5 %  |
| Egodefens . . . .         | 1.98       | 2.54     | 1.71           | 38 %   | 4 %  |
| Behovspersistens          | 1.75       | 2.39     | 1.61           | 36 %   | 3 %  |

Tabell 11 antyder at testen ikke gir særlig pålitelige resultater. Hvor strenge krav en skal stille i denne retning, vil selvsagt i høy grad være bestemt av hvilke spesielle formål en har for øye. Taylors konklusjon — at testen «er uskikket til vurdering av enkeltindivider» — synes stort sett å være holdbar også for norske forhold sett fra en rent psykometrisk synsvinkel. Spørsmålet om dens kliniske anvendbarhet er ikke dermed besvart. Heller ikke spørsmålet om dens anvendbarhet i eksperimentell psykologisk forskning, hvor konsentrasjon om grupper mer enn enkeltindivider er fremherskende. I denne forbindelse vil det ofte i første rekke være de enkelte testkategoriers evne til å differensiere på en statistisk signifikant måte som er avgjørende.

Testens tilsynelatende høyst begrensede pålitelighet må utvilsomt til en viss grad tilskrives det meget ensartede utvalg fpp. som ble anvendt. Innenfor mer avvikende, uensartede utvalg er det rimelig å anta at testens reliabilitet vil stige. Men en kan vanskelig komme bort fra at testen også har visse indre svakheter.

En svakhet som Taylor sterkt påpeker, er testens relativt lille antall ledd, sett i forhold til at den skal gi data til hele 6 variabler. Og videre at enkelte av leddene sterkt synes å favorisere svar i én skåringskategori fremfor andre, noe som medfører en ytterligere aksentuering av denne svakhet.

For å få en oversikt over i hvilken grad en slik favorisering virkelig kan tilskrives testen og ikke særegenheter ved det spesielle utvalg

TABELL 12

*Prosentvis fordeling av skårer på forskjellige kategorier og ledd*

| Ledd<br>nr. | Kategorier |      |      |      |      |      |
|-------------|------------|------|------|------|------|------|
|             | Ex         | In   | Im   | H-D  | E-D  | B-P  |
| 1           | 41.2       | 2.9  | 55.9 | 4.4  | 86.8 | 8.8  |
| 2           | 11.8       | 80.1 | 8.1  | 11.8 | 49.2 | 39.0 |
| 3           | 85.3       | 8.8  | 5.9  | 25.7 | 33.1 | 41.2 |
| 4           | 17.7       | 2.9  | 79.4 | 14.7 | 43.4 | 41.9 |
| 5           | 18.4       | 80.9 | 0.7  | 20.6 | 27.2 | 52.2 |
| 6           | 39.0       | 44.1 | 16.9 | 11.0 | 39.7 | 49.3 |
| 7           | 82.4       | 17.6 | 0    | 0    | 99.3 | 0.7  |
| 8           | 44.9       | 2.9  | 52.2 | 39.0 | 58.8 | 2.2  |
| 9           | 86.0       | 3.7  | 10.3 | 8.8  | 16.2 | 75.0 |
| 10          | 91.9       | 6.6  | 1.5  | 0.7  | 97.8 | 1.5  |
| 11          | 72.8       | 0    | 27.2 | 2.9  | 94.1 | 3.0  |
| 12          | 33.8       | 26.5 | 39.7 | 5.9  | 21.3 | 72.8 |
| 13          | 89.0       | 6.6  | 4.4  | 6.6  | 30.1 | 63.3 |
| 14          | 64.0       | 0.7  | 35.3 | 3.7  | 80.1 | 16.2 |
| 15          | 14.7       | 5.2  | 80.1 | 6.6  | 69.1 | 24.3 |
| 16          | 79.4       | 20.6 | 0    | 0    | 57.4 | 42.6 |
| 17          | 26.5       | 67.6 | 5.9  | 7.4  | 74.2 | 18.4 |
| 18          | 33.1       | 14.7 | 52.2 | 22.1 | 20.6 | 57.3 |
| 19          | 23.5       | 76.5 | 0    | 0.7  | 99.3 | 0    |
| 20          | 30.1       | 24.3 | 45.6 | 20.6 | 78.7 | 0.7  |
| 21          | 23.5       | 75.0 | 0.5  | 53.7 | 40.4 | 5.9  |
| 22          | 41.2       | 13.2 | 45.6 | 73.5 | 24.3 | 2.2  |
| 23          | 86.8       | 0.7  | 12.5 | 25.7 | 50.0 | 24.3 |
| 24          | 16.9       | 1.5  | 81.6 | 52.9 | 39.0 | 8.1  |
| M           | 48.1       | 24.3 | 27.6 | 17.5 | 55.4 | 27.1 |

Taylor anvendte, foretok også vi en spesiell leddanalyse av vårt materiale.

Tabell 12 gir en oversikt over resultatene.

Tabellen viser at det også hos oss foreligger en påtagelig skjevhet i skårefordelingen ved mange av leddene.

Høyest kategorifavorisering finner vi for E-D på ledd nr. 7 og 19 med hele 99 %, lavest for Im på ledd nr. 7, 16 og 19, for H-D på ledd nr. 7 og 16, for B-P på ledd nr. 19, og for In på ledd nr. 11, alle med 0 %.

En slik uttalt kategorifavorisering ved mange av leddene gir utvilsomt en viss forklaring på den lave indre konsistens, men den gir også en antydning om mulige fremgangsmåter for å gjøre den indre konsistens høyere. En vei å gå vil være å utelate de ledd som virker sterkest favoriserende, eller skifte dem ut med nye ledd med en mer allsidig differensieringsevne. De ledd det her i første rekke er tale om, er ledd nr. 7, 10 og 19. Da testens lengde allerede på forhånd er sterkt begrenset, er det siste alternativ, en samtidig tilføyelse av nye ledd, den mest fordelaktige løsning. Muligens vil det vise seg ønskelig ikke bare å skifte ut ledd, men kanskje også å gjøre selve testen noe lenger, f.eks. utvide den fra 24 til 28 eller 32 ledd.

## SAMMENLIGNING

### MED AMERIKANSKE KATEGORINORMER

Vi har i behandlingen av vårt empiriske materiale hovedsakelig konsentrert oss om de forskjellige skåringskategorier. Ett av de første spørsmål som ble reist, var forholdet mellom våre gjennomsnittsverdier og de amerikanske kategorinormer for menn.

Det er særlig resultatene fra tre amerikanske undersøkelser det er av interesse å foreta en sammenligning med, nemlig Taylors, Rosenzweigs og Bernards resultater. I de to første tilfellene er resultatene blitt fremstilt i form av aritmetiske gjennomsnitt og standardavvik. Tabell 13 gir en oversikt over resultatene av disse to undersøkelser, sammen med våre egne data.

En sammenligning mellom våre og Taylors resultater viser signifikante forskjeller for alle kategorier unntatt hindringsdominans. Vårt utvalg er mer ekstrapunitivt og sterkere preget av ego-defensive reaksjonsformer, mens Taylors ligger høyere i forbindelse med intro-punitiv, impunitiv og behovspersistente tendenser.

Det er vanskelig å gi noen entydig forklaring på forskjellene. Taylors utvalg består av 260 voksne amerikanske menn testet i forbindelse med personundersøkelse eller yrkesrettleiing. Vårt utvalg av 136 offisersaspiranter, alle med gymnasium som utdanningsmessig bakgrunn. Det er ikke urimelig å anta vesentlige forskjeller mellom de to utvalg både med hensyn til alder og formell skolegang.

Når det gjelder nettopp disse faktorer, er sannsynligvis Rosenzweigs utvalg mer i overensstemmelse med vårt eget (kfr. s. 22). Forskjellene er da også her av til dels betydelig mindre størrelsesorden. Vårt utvalg skårer noe høyere på den ekstrapunitiv og ego-defensive kategori,

TABELL 13

Sammenligning mellom eget materiale og Taylors og Rosenzweigs normer for menn

| Kategori | Eget utvalg<br>(N=136) |          | Taylors utv.<br>(N=260) |          | Rosenzweigs utv.<br>(N=236) |          | Eget vs.<br>Taylors<br>utv. | Eget vs.<br>Rosen-<br>zweigs utv. |
|----------|------------------------|----------|-------------------------|----------|-----------------------------|----------|-----------------------------|-----------------------------------|
|          | M                      | $\sigma$ | M                       | $\sigma$ | M                           | $\sigma$ | CR                          | CR                                |
| Ex       | 11.54                  | 2.91     | 8.62                    | 2.65     | 10.80                       | 3.19     | 9.73                        | 3.96                              |
| In       | 5.84                   | 1.70     | 7.29                    | 1.70     | 6.72                        | 1.99     | 8.06                        | 6.09                              |
| Im       | 6.63                   | 2.21     | 7.79                    | 2.07     | 6.48                        | 2.28     | 5.04                        | —                                 |
| H-D      | 4.19                   | 1.89     | 4.19                    | 1.53     | 4.80                        | 1.87     | —                           | 4.13                              |
| E-D      | 13.30                  | 2.54     | 11.43                   | 2.26     | 12.72                       | 2.71     | 7.19                        | 3.34                              |
| B-P      | 6.51                   | 2.39     | 8.09                    | 2.39     | 6.48                        | 2.47     | 6.32                        | —                                 |

mens Rosenzweigs har noe sterkere intro-punitive og hindringsdominante tendenser. Ikke i forhold til noen av kategoriene er imidlertid forskjellen i gjennomsnittsverdier (i prosent) over 4%.

Også Bernards normer for menn (6, s. 328) gir et bedre utgangspunkt for sammenligning enn Taylors utvalg. Bernards utvalg består av 92 menn, hovedsakelig studenter. En ulempe ved Bernards normer er imidlertid at de alle er uttrykt i median- og kvartilverdier, noe som i høy grad vanskeliggjør en mer eksakt beregning av forskjeller. Tabell 14 gir en oversikt over våre egne resultater uttrykt i slike målenheter.

TABELL 14

Egne empiriske resultater uttrykt i median- og kvartilverdier

| Retnings-<br>kategorier          | Formkategorier |       |       |       |       |       |       |       |       | Total for<br>retn.kateg. |       |       |
|----------------------------------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|--------------------------|-------|-------|
|                                  | H-D            |       |       | E-D   |       |       | B-P   |       |       | $Q_1$                    | $Mdn$ | $Q_3$ |
|                                  | $Q_1$          | $Mdn$ | $Q_3$ | $Q_1$ | $Mdn$ | $Q_3$ | $Q_1$ | $Mdn$ | $Q_3$ |                          |       |       |
| Ex                               | 0.6            | 1.6   | 3.0   | 4.6   | 7.0   | 8.8   | 1.8   | 2.6   | 3.9   | 9.7                      | 12.0  | 13.6  |
| In                               | 0              | 0.9   | 1.4   | 2.1   | 3.1   | 4.0   | 0.8   | 1.7   | 2.7   | 4.8                      | 5.8   | 6.9   |
| Im                               | 0.6            | 1.2   | 2.1   | 2.2   | 3.2   | 4.7   | 0.9   | 1.6   | 2.5   | 5.1                      | 6.4   | 8.2   |
| Total for<br>formkate-<br>gorier | 2.9            | 4.1   | 5.4   | 11.7  | 13.3  | 15.0  | 4.8   | 6.4   | 8.1   |                          |       |       |

Overensstemmelsen mellom våre og Bernards resultater synes å være ganske stor. Bernards utvalg ligger noe høyere med hensyn til hindringsdominans, ego-defensive og intropunitive tendenser, vårt utvalg noe høyere med hensyn til behovspersistens, ekstrapunitive og impunitive tendenser, men medianene i de to utvalg divergerer ikke i forhold til noen av kategoriene med over 2 %.

Dette må sies å antyde relativt små forskjeller, særlig sett på bakgrunn av at gjennomsnittsverdiene for behovspersistens og ego-defens, intropunitive og ekstrapunitive tendenser, i vårt og Taylors utvalg divergerer med henholdsvis 7 og 8, 6 og 12 %.

Det er naturligvis ytterst vanskelig å si hvilke forhold som er utslagsgivende, når det gjelder de tilsynelatende små forskjeller mellom våre og Bernards, og de relativt store forskjeller mellom våre og Taylors resultater.

Våre resultatets større overensstemmelse med Rosenzweigs enn med Taylors bringer umiddelbart tanken inn på at alders- og utdannelsesforhold er av betydning, mens den tilsynelatende enda større likhet med Bernards resultater gjør det nødvendig også å søke etter andre forklaringer.

I det store og hele synes administrasjonsformen i Taylors og Rosenzweigs undersøkelser å ha vært nokså lik, mens derimot Bernard ved sin bruk av testen som en selvadministrert prøve klart skiller seg ut fra begge de to andre.

Også en annen forskjell er muligens av betydning. En ikke ubetydelig del av Bernards fpp. ble prøvd som arbeidssøkere, og B-F-testen ble nyttet som et ledd i et større utvalgsbatteri. «Samnsynligvis var innstillingen blant mange, hvis ikke blant alle disse fpp.,» skriver Bernard — «å svare på en slik måte at det ikke ville hindre dem i å få det arbeid de søkte» (7, s. 333).

Bernard skriver videre at han ved spørsmål etter prøvingen forsøkte å klarlegge hvilket svarnivå som var blitt nyttet. Flere ga da uttrykk for at de hadde gitt svarene sine ut fra slik de ville like å handle. En kan spørre seg hvorvidt B-F-testen overhodet kan ha noen verdi i en slik sammenheng. «Til tross for de forskjelligartede innstillinger blant fpp.» hevder imidlertid Bernard at han «har funnet at B-F-testen brukt som selvadministrert prøve gir verdifullt materiale til vurdering av visse personlighetstrekk og til forutsigelse av bestemte atferdsmønstre» (7, s. 333).

Våre egne resultater viser en tilsynelatende størst overensstemmelse med Bernards resultater, til tross for at vår administrasjonsform sannsynligvis ligger betydelig nærmere Rosenzweigs og Taylors enn Bernards.

Vårt utvalg består av en bestemt type studenter. Det er ikke utenkelig at offisersaspiranter viser merkbart forskjellige reaksjonstendenser enn studenter med andre utdannelsesplaner, at offisersaspiranter viser tendenser nettopp i de retninger som f.eks. styrkes ved å gi testen som selvadministrert prøve. Det er imidlertid et annet aspekt vi her

i første rekke vil trekke frem. Når det gjelder selve prøvesituasjonen, er det høyst sannsynlig at det er flest likhetspunkter mellom vår og Bernards undersøkelse.

Fpp. i vår undersøkelse var alle preget av ønske om en bestemt utdannelse, men samtidig en viss usikkerhet om hvorvidt den ville la seg realisere. Skjønt alle våre fpp. ble forsikret om at svarene de ga, ikke ville ha noen betydning for deres utdannelse som offiserer, er det allikevel rimelig å anta at en høy grad av omtenkksomhet og refleksjon gjorde seg gjeldende. I denne retning peker bl.a. den relativt lange prøvetid som gjennomgående viste seg nødvendig.

En høy grad av refleksjon må vi også anta var til stede i den delen av Bernards materiale hvor B-F-testen ble anvendt som utvalgsinstrument. Bruken av testen som selvadministrert prøve har dessuten trolig mer fremmet enn hemmet en slik tendens.

Rent generelt skulle vi vente at vårt og Bernards utvalg har gitt relativt mer opinionsorienterte svar, at de stort sett har vært mer kontrollerte, enn utvalg som har fått testen på et mer 'frivillig' grunnlag.

Den mest påfallende forskjell mellom vårt og de andre utvalg er selvsagt deres forskjellige nasjonale bakgrunn. Det er rimelig å anta nasjonale forskjeller i reaksjonstendenser på frustrasjon. Men nøyaktig på hvilken måte og i hvilken grad disse forskjeller spiller inn, er det ytterst vanskelig å ha noen sikker mening om.

Vi vet ikke i dag hvorvidt det foreligger forskjeller mellom forskjellige utvalg av studenter, hvorvidt f.eks. norske offisersaspiranter reagerer forskjellig fra andre studenter. Heller ikke vet vi med sikkerhet hvilken betydning som bør tillegges testens administrasjonsform, om f.eks. det er slik at selvadministrering fremkaller mer kontrollerte svar enn gruppe- eller individualadministrering, — eller hvilket betydning selve testsituasjonen har i det enkelte tilfelle, om det er ønskelig til enhver tid å skille mellom testsituasjoner av «frivillig» og «ufrivillig» karakter.

Vi har ovenfor tatt utgangspunkt i forskjellene mellom gjennomsnittsverdiene i vår og andres undersøkelser. Det har vist seg å være relativt lett å finne mulig forklaring på forskjellene — selv uten å anta nasjonale forskjeller, og stort sett er heller ikke forskjellene større enn at de amerikanske testnormene skule kunne anvendes under norske forhold, under forutsetning av at de anvendes på en skjønnsom måte og med tilbørlig hensyntagen til særegenheter ved det enkelte tilfelle. Dette må ikke forstås slik at det ikke foreligger behov for testnormer basert på mer representative norske utvalg enn vårt, men i mangel av slike synes de amerikanske normer rent foreløpig å kunne gi visse holdepunkter — særlig når en ikke innskrenker seg til normer basert på en enkelt undersøkelse, men forsøker å se det hele i en større sammenheng. I denne forbindelse kan utvilsomt også våre egne data ha en viss praktisk verdi.

## SAMMENLIGNING MED SKOLEHJEMSGUTTERS REAKSJONSTENDENSER

En sammenligning mellom utvalg som har den samme nasjonale bakgrunn, byr vanligvis på betydelige fordeler når det gjelder kontrollerbarhet. Også mulighetene for å fremsette prøvbare hypoteser om forskjeller og likheter mellom utvalg vil være større.

Bruken av B-F-testen i Norge har hittil vært sterkt begrenset. Ved siden av vår undersøkelse foreligger bare en undersøkelse av Brun-Gulbrandsen (10) over elevene ved Buskerud skolehjem. En sammenligning av resultatene i de to undersøkelser kan være av interesse.

Skolehjemsgutter har praktisk talt alltid gjort seg skyldig i asosiale handlinger. Selve deres anbringelse på skolehjem gir i alminnelighet uttrykk for en mangel på kontroll av aggressive tendenser. Ved siden av skolehjemsutvalgets høyere ekstrapunitive tendenser skulle vi også vente at dette utvalg hadde liten tilpasningsevne, dvs. at de gjennomgående skåret lavt på den behovspersistente kategori. Tabell 15 gir en oversikt over våre egne og Brun-Gulbrandsens resultater uttrykt i prosentverdier.

TABELL 15

*Sammenligning mellom offisersaspiranter og skolehjemsgutters testresultater*

| Kategori | Eget utvalg<br>(N=136) |      | Brun-Gulbrandsens<br>utvalg (N=27) |      | Diff. | CR * | p     |
|----------|------------------------|------|------------------------------------|------|-------|------|-------|
|          | M                      | σ    | M                                  | σ    |       |      |       |
| Ex       | 48.1                   | 12.1 | 56.5                               | 15.5 | -8.4  | 2.66 | .01   |
| In       | 24.3                   | 7.1  | 22.8                               | 10.0 | 1.5   | 0.77 |       |
| Im       | 27.6                   | 9.2  | 20.7                               | 9.9  | 6.9   | 3.39 | .01   |
| H-D      | 17.5                   | 7.9  | 22.4                               | 9.2  | -4.9  | 2.61 | .01   |
| E-D      | 55.4                   | 10.6 | 56.2                               | 12.5 | -0.8  | 0.29 |       |
| B-P      | 27.1                   | 10.0 | 21.4                               | 10.9 | 5.7   | 2.52 | ≈ .01 |

\* En har antatt differensene normalfordelt.

En merker seg umiddelbart skolehjemsutvalgets høyere ekstrapunitive og hindringsdominante tendenser, og vårt eget utvalgs høyere impunitive og behovspersistente tendenser. Resultatene må sies nøye å svare til hva vi ut fra teoretiske betraktninger skulle vente å finne.

På ny blir en selvsagt stilt overfor spørsmålet om hvilken innvirk-



ning de to utvalgs forskjellige utdannelse og alder kan ha hatt på resultatene.

Ifølge en amerikansk undersøkelse (kfr. s. 44) synes det bl.a. å foreligge en viss positiv korrelasjon mellom generell intelligens og hindringsdominante og behovspersistente svartendenser. Skolehjemmutvalgets gjennomsnittlige intelligens (ved Terman-Merrill-prøven) var  $IQ = 83$ . Utvalgets høyere hindringsdominans antyder derfor at det er andre faktorer enn intelligens som i dette tilfelle er utslagsgivende.

Indirekte blir våre resultater støttet av en undersøkelse av Gatling (24). Ved bruk av en modifisert utgave av Rosenzweigs opprinnelige frustrasjonstest (kfr. s. 11) foretar Gatling en systematisk sammenligning av reaksjonsretningen blant kriminelle og ikke-kriminelle barn. Han finner at de kriminelle barna skårer signifikant mer ekstrapunitivt, de ikke-kriminelle signifikant mer intro-punitivt. Av særlig interesse er Gatlings resultater fordi aldersfaktoren her er konstant.

Forskjellene mellom vårt og Brun-Gulbrandsens utvalg viser seg på én unntagelse nær å gå i samme retning som forskjellene mellom Frys utvalg av mannlige studenter og fanger (kfr. s. 42 f.). Mens Fry fant at utvalget med ikke-kriminelle skåret høyere enn kriminelle på den ekstrapunitive og lavere på den intro-punitive kategori, er forholdet hos oss det motsatte.

Dette reiser faktisk spørsmålet om voksne kriminelle, fengselsfanger, skiller seg ut på en signifikant måte fra skolehjemselever. Det er unektelig en interessant hypotese at det i første rekke er de mest intro-punitive skolehjemselever som faller tilbake og senere blir rekruttert som fengselsklientel. En mulighet er at en del av de unge kriminelle, gjennom de offentlige reaksjoner, læres opp til å reagere verbalt intro-punitivt, uten samtidig i nevneverdig grad å modifisere sine dypere ekstrapunitive tendenser, og at det nettopp er disse som senere faller tilbake.<sup>23</sup> Vi står her overfor sentrale behandlingsmessige spørsmål i forbindelse med unge lovovertredere. Ennå i dag er behovet for dybdepsykologisk behandling på langt nær fullt erkjent.

Hypotesen ovenfor har en viss meget interessant parallell i Albee og Goldmans resultater: Blant psykiatriske pasienter viser de som skårer høyt på den intro-punitive kategori, men for øvrig er preget av ekstrapunitive tendenser, en særdeles dårlig prognose (kfr. s. 63 og s. 76).

## UNDERSØKELSE OVER KORRELASJONEN MED TILPASSINGSVURDERINGER

Av det totale antall brukbare testjournaler i vår undersøkelse, 136, var 79 besvart av aspiranter ved Krigsskolen. For hver av disse forelå det ved prøvingen et relativt omfattende intervju- og observasjonsmateriale — innsamlet av psykologer ved *Forsvarets psykologiske av-*

<sup>23</sup> Lindzey og Goldwyns undersøkelse med barneutgaven av testen, hvor de fant at skolehjemsbarn som tilpasset seg skolehjemssituasjonens krav, skåret mer intro-punitivt enn andre, kan f.eks. gi uttrykk for en slik overfladisk læringsprosess (kfr. s. 42).

*deling*. I den psykologiske undersøkelse av hver enkelt aspirant inngikk også en kvantitativ tilpassingsvurdering. Denne ble foretatt i forhold til to 9-delte vurderingsskalaer: En for sosial tilpassing — «en personlig vurdering av hvorledes aspiranten vil tilpasse seg sosialt under de militære skoleforhold han evt. skal underkastes», og én for generell tilpassing under sivile forhold — bygd på «inntrykk av hvorledes aspiranter *har* tilpasset seg de forhold han har levd under, med særlig vekt på de siste år».

Stort sett var de to tilpassingsvurderinger for den enkelte den samme. I begge tilfelle la man stor vekt på en rent personlighetsmessig vurdering — en vurdering av aspirantenes truselorientering *kontra* målorientering, eller i vår terminologi — deres tilbøyelighet til ego-defensive *kontra* behovspersistente reaksjonsformer.

Med dette som bakgrunn fant vi det av interesse å foreta en beregning av korrelasjonen mellom tilpassingsvurdering og B-F-testens skåringskategorier.

Et samlet mål på tilpassingsvurdering fikk vi ved å beregne hver enkelt aspirants gjennomsnittsskåre på de to skalaer. Tilpassingsvurderingene er på ingen måte fri for feilkilder. Noe pålitelig kriterium på grad av målorientering kan de neppe sies å representere. Deres kvantitative spredning i vårt utvalg var dessuten meget begrenset. Noen særlig høy korrelasjon gjorde vi ikke regning med å finne, mens vi på den annen side mente det ville være berettiget å reise en del tvil om B-F-testens validitet hvis det overhodet ikke skulle vise seg å foreligge noen korrelasjon.

Tabell 16 gir en oversikt over våre resultater.

TABELL 16

*Korrelasjonen (r) mellom tilpassingsvurdering og B-F-testens kategorier og dens signifikansnivå (p)*

| Retn.kategorier | r    | p | Formkategorier | r    | p   |
|-----------------|------|---|----------------|------|-----|
| Ex              | -.04 | — | H-D            | .02  | —   |
| In              | .01  | — | E-D            | -.19 | .08 |
| Im              | .05  | — | B-P            | .20  | .07 |

Tabellen viser at det ikke foreligger noen korrelasjon mellom tilpassingsvurdering og 4 av B-F-testens kategorier. Derimot synes det å foreligge en viss korrelasjon mellom tilpassingsvurdering og E-D- og B-P-kategoriene. Heller ikke disse to korrelasjoner er særlig høye, men de går begge i ventet retning. At vi finner en viss korrelasjon nettopp med disse kategorier, er for øvrig ganske forståelig på bakgrunn av tilpassingsvurderingenes karakter.

En nærmere analyse av de enkelte skåringsfaktorer viser at E og m er de faktorer som differensierer best mellom god og dårlig «tilpassing». Også i andre undersøkelser har m-faktoren blitt påvist å ha størst differensierende evne (kfr. s. 49).

Når det gjelder å bedømme B-F-testens anvendbarhet, kan man med en viss rett hevde at det ikke er korrelasjonen mellom hver enkelt kategori og tilpassingsvurderingene som er av betydning, men deres multiple korrelasjon med dette kriterium.

Ved utvalg av piloter i det amerikanske flyvåpen under siste verdenskrig ble det f.eks. anvendt et testbatteri på hele 18 tester. Utvalg basert på hver enkelt av testene ga resultater som bare var litt bedre enn tilfeldige. Når de 18 testene ble satt sammen i et batteri, ble det derimot oppnådd en validitet på  $+ .60$  (1, s. 259).

Vi har begrenset oss til en beregning av den multiple korrelasjon mellom E-D- og B-P-kategoriene og tilpassingsvurderingene. Korrelasjonen mellom E-D og B-P gir en  $r = -.71$ , og den multiple korrelasjon gir  $r_{1,23} = .21$  og  $\sigma_{r_{1,23}} = .088$ .<sup>24</sup> Dette gir videre  $CR = 2.39$ , hvilket vil si at den multiple korrelasjonskoeffisient er signifikant på ca. 2 % nivået.

Vi står m.a.o. overfor en relativt pålitelig korrelasjon hvis vi tar hensyn til fpp.s skåreverdier både på E-D- og B-P-kategoriene, men korrelasjonen er likevel så lav at forutsigelser basert på B-F-testen alene er meget usikre.

Det er ikke urimelig å anta at testen ville ha vist større differensieringsevne, hadde utvalget med fpp. vært mer uensartet. Det er imidlertid et annet forhold vi her i første rekke vil trekke frem.

Hvis vi tenker oss at testen og tilpassingsvurderingene registrerer forskjellig atferdsnivåer, vil den relativt lave korrelasjon mellom dem ikke være til hinder for at de begge hver for seg er valide. Hvis tilpassingsvurderingene f. eks. i første rekke gir uttrykk for eksplisitte ego-defensive tendenser og B-F-testen hovedsakelig har sin forankring på implisitt eller latent atferdsnivå, skulle vi faktisk vente å finne en relativt begrenset korrelasjon.

Det foreligger ingen sikre holdepunkter til bedømmelse av tilpassingsvurderingenes nivåforankring. En kan ikke se bort fra at de muligens er noe inkonsistente, bestemt av de enkelte uttakspsykologers forskjelligartete trening og erfaring, men det er allikevel nærliggende å anta en hovedsaklig eksplisitt nivåforankring.

Når det gjelder B-F-testens nivåforankring, er vi også på noe usikker grunn. Vi har tidligere diskutert dette spørsmål relativt inngående. Vi hevdet da at B-F-testen sannsynligvis har en høyst labil nivåforankring, og at det er ønskelig å få denne spesielt klarlagt i det enkelte tilfelle. Uten en slik klarlegging, hevdet vi, må B-F-testen sies å ha en svært begrenset nytteverdi.

---

<sup>24</sup> I formelen for den multiple korrelasjon blir vektallene  $B_2 = -.04$  og  $B_3 = .05$ .

## UNDERSØKELSE OVER TESTENS NIVÅFORANKRING

Spørsmålet om svarnivå er av sentral betydning når en skal vurdere resultater på B-F-testen. Noen utprøvet fremgangsmåte for registrering av svarnivåer foreligger imidlertid ennå ikke.

Det ble i den foreliggende undersøkelse foretatt et første forsøk på å utarbeide en slik fremgangsmåte.

B-F-testen ble som nevnt administrert som gruppeprøve. Umiddelbart etter besvarelsen av selve prøven, ble samtlige fpp. forelagt et kort spørreskjema hvor de ble bedt om å angi på hvilket grunnlag de hadde gitt svarene sine.

Følgende spørsmål ble stilt:

- A. Da du skrev replikker, tenkte du deg noen gang selv i samme situasjon og hva du ville ha sagt i situasjonen?
- B. Da du skrev replikker, svarte du noen gang det du antok du selv *faktisk ville* ha sagt i samme situasjon?
- C. Da du skrev replikker, svarte du noen gang det du mente du *burde* ha sagt i samme situasjon, uten tanke på hvorledes du faktisk mest sannsynlig ville ha oppført deg i situasjonen?

Etter hvert av spørsmålene var angitt en 5-delt skala: *alltid, ofte, iblant, sjelden, aldri*, og fpp. ble anmodet om å angi det svar de syntes passet. (Kfr. Appendiks I.)

Hvert av spørsmålene tok sikte på å gi opplysninger om i hvilken grad de forskjellige svarnivåer var blitt nyttet, med spørsmål A i hvilken grad fpp. hadde latt ubevisste, ikke-jegkontrollerte impulser få slippe til, med spørsmål B i hvilken grad svarene ga uttrykk for vanlige jeg-kontroll, og med spørsmål C i hvilken grad super-kontroll var til stede.

Ved å gi svarene på hvert av spørsmålene vekt tall fra 1 til 5, fra liten til stor «ladning», håpet vi å finne frem til hver enkel fpp.s sannsynlige svarnivå.

Når det gjelder den anvendte fremgangsmåte for nivåbestemmelse, er det en innvending som er nærliggende. Hvilken garanti har man for at fpp. ikke rasjonaliserer, at de faktisk er oppmerksomme på hvilket nivå de svarte på, og at de svarer sant på spørsmålene som blir stilt dem?

Spørsmålene er slik konstruert at de gir mulighet for en viss kontroll når det gjelder logikken i fpp.s svar. Det lar seg gjøre å undersøke om fpp.s svar er innbyrdes holdbare, om høy «ladning» på ett nivå f.eks. blir fulgt av lav på de to andre.

Nå er det imidlertid ikke slik at bristende logikk øyeblikkelig er tegn på usannhet eller rasjonalisering. Rasjonalisering kan ofte finne sted med en plettfri logikk.

Den anvendte fremgangsmåte gir ingen helt sikre opplysninger med hensyn til svarnivå, men visse verdifulle antydninger. Metoden har selv

behov for uavhengig validering. Vi vil her imidlertid som arbeidshypotese gå ut fra at den er valid.

TABELL 17

*Oversikt over samtlige fpp.s «ladninger» på de forskjellige nivåer*

| Spm. | Antatt svarnivå | 1  | 2  | 3  | 4  | 5  | M    | $\sigma$ |
|------|-----------------|----|----|----|----|----|------|----------|
| A    | implisitt       | 30 | 89 | 43 | 3  | 1  | 2.13 | 0.74     |
| B    | eksplisitt      | 2  | 6  | 47 | 82 | 29 | 3.78 | 0.81     |
| C    | opinion         | 36 | 44 | 56 | 25 | 5  | 2.51 | 1.08     |

Tabell 17 gir en oversikt over fordelingen av svar på de forskjellige spørsmål. Tabellen antyder at eksplisitt svarnivå er det nivå som er blitt nyttet hyppigst. Men tabellen viser også at det foreligger en betydelig spredning. Dermed støttes også empirisk holdbarheten av vår tidligere antagelse med hensyn til B-F-testens nivåabilitet.

Et relativt lite antall fpp. synes å ha nyttet eksplisitt svarnivå gjennom hele testen. Mange synes iblant å ha nyttet opinionsnivå og implisitt nivå. Og enkelte synes endog å ha nyttet disse svarnivåer gjennom hele testen.

En bemerkelsesverdig side ved resultatene er at implisitt svarnivå synes å ha blitt svært lite nyttet. Noen virkelig proaktiv test i den forstand at den gir uttrykk for upersonlige, ubevisste impulser og tendenser, synes testen bare å ha vært for et meget lite antall.

Under diskusjonen av våre testresultaters overensstemmelse med amerikanske kategorinormer fant vi det bl. a. ønskelig å anta at våre fpp. i en viss grad hadde vært influert av ideelle overveielser. Den situasjon de befant seg i ved testingen, usikkerheten med hensyn til hvorvidt deres utdannelsesplaner ville la seg realisere, antok vi virket befordrende på jeg-kontrollerte svar. Med et slikt utgangspunkt fant vi det høyst naturlig at våre resultater viste størst samsvar med Bernards testnormer.

Vår antagelse at våre fpp. i høy grad var influert av jegkontroll, synes altså å bli støttet. I hvilken grad dette avspeiler deres spesielle testsituasjon og ikke er uttrykk for begrensninger ved selve testen, er det umulig å ha noen sikker mening om. Ut fra den spesielle test-situasjon fpp. befant seg i, kunne en kanskje ha ventet en enda høyere grad av opinionsorienterte svar.

Ved siden av inter-individuelle forskjeller antydes også en ikke ubetydelig intra-individuell variasjon: en og samme person synes ofte å reagere på forskjellige nivåer på forskjellige deler av testen.

Det er en svakhet ved metoden at den ikke spesifiserer dette mer nøyaktig, slik at det post facto er mulig å foreta en separat analyse av den enkelte fpp.s reaksjonstendenser på forskjellige nivåer.

En kan spørre seg hvilke faktorer det er som bestemmer en persons valg av svarnivå, hva som er karakteristisk for personer som velger henholdsvis implisitt, eksplisitt og opinionsnivå. Det er ikke utenkelig at kjennskap til en persons valg i denne forbindelse i seg selv kan være av betydelig diagnostisk interesse.

## UNDERSØKELSE OVER TESTENS VALIDITET

Validitetsundersøkelser av projektive prøver reiser meget kompliserte problemer. Særlig når det gjelder å finne egnete kriterier. Å validere en prøve ved å beregne korrelasjonen med en annen prøve som selv ikke er påvist å være valid, vil selvsagt ikke kunne gi grunnlag for sikre slutninger. Heller ikke observasjonsdata er alltid egnet. De vil lett avspeile den spesielle situasjon observasjonene har funnet sted under, og de vil lett også være influert av observatørens egen personlighetsstruktur. Men den mest vesentlige innvending som kan rettes mot denne form for validering, har sitt utgangspunkt i nivåproblematikken. Forsøk på å validere en projektiv prøve som har som hovedformål å registrere fantasiliv og ubevisste prosesser, mot manifest observert atferd, strider mot testens grunnleggende forutsetninger.

En beregning av korrelasjonen mellom tilpassingsvurderinger og B-F-testens kategorier gir oss derfor i liten grad mål på testens «validitet» medmindre vi har kjennskap til at svarnivået på testen tilsvarer tilpassingsvurderingenes nivåforankring.

Hvis vi antar at tilpassingsvurderingene hovedsaklig er forankret på eksplisitt nivå, følger at vi skulle vente en større korrelasjon med testen, jo mer testsvarene er gitt på dette nivå. Ved å konsentrere oss om det utvalg blant våre fpp. som vi har grunn til å anta har gjort konsistent bruk av eksplisitt nivå på testen, skulle vi således få et visst grunnlag for beregning av korrelasjoner til belysning av testens «validitet».

Det første problem som møtte oss når vi skulle ta fatt på en slik analyse, var hvordan vi mest hensiktsmessig skulle klassifisere vårt materiale med hensyn til svarnivå. Rent prinsipielt er det mulig å skille mellom en relativ og en mer absolutt klassifiseringsmetode.

Vi har gjort bruk av begge deler. Vår første fremgangsmåte var av mer absolutt karakter, idet vi la vekt på å differensiere mellom mer og mindre konsistent bruk av de forskjellige nivåer. Skjema 4 illustrerer klassifiseringsmåten.

Skjemaet bygger på den enkelte f.p.s svar på alle tre spørsmål. A står for «ladning» på implisitt nivå, B for «ladning» på eksplisitt nivå og C for «ladning» på opinionsnivå. Med «ladning» siktes til i hvilken grad f.p. angir å ha nyttet det spesielle nivå — om vedkommende har gjort bruk av det alltid, ofte, iblant, sjelden eller aldri.

Første trinn i klassifiseringen ble foretatt i forhold til tendenser. Alle som viste seg å ha en «ladning» på et spesifikt svarnivå som var lik eller større enn deres «ladning» på de to andre nivåer, ble definatorisk tillagt en *tendens* på dette nivå. Personer som hadde samme

## SKJEMA 4

### *Metode for klassifisering av fpp. etter grad av konsistens*

| Nivå                 | Tendens           | Overvekt    | Konsistens  |
|----------------------|-------------------|-------------|-------------|
| Implisitt . . . . .  | $B \leq A \geq C$ | $B < A > C$ | $A > B + C$ |
| Eksplisitt . . . . . | $A \leq B \geq C$ | $A < B > C$ | $B > A + C$ |
| Opinion . . . . .    | $A \leq C \geq B$ | $A < C > B$ | $C > A + B$ |

«ladning» på alle nivåer, ble med andre ord tillagt en tendens i forhold til alle nivåer, personer som hadde den samme «ladning» på to nivåer og en mindre «ladning» på det tredje nivå, ble tillagt en tendens i forhold til de to første, osv.

Neste trinn besto i å skille ut de som ikke bare hadde en tendens i forhold til spesifikt nivå, men også en virkelig overvekt på nivået. *Overvekt* ble tillagt dem som hadde en høyere «ladning» på ett nivå enn på de to andre nivåer.

Blant dem som hadde en slik overvekt, ble igjen skilt ut de som faktisk viste en høy grad av konsistens. *Konsistens* ble tillagt dem som hadde en «ladning» på ett nivå som var større enn summen av «ladningene» på de to andre nivåer.

En klassifisering av våre fpp. — i alt 79 — etter disse linjer er angitt i tabell 18.

### TABELL 18

#### *Fordeelingen av fpp. etter grad av konsistens*

| Nivå                 | Tendens | Overvekt | Konsistens |
|----------------------|---------|----------|------------|
| Implisitt . . . . .  | 14      | 1        | —          |
| Eksplisitt . . . . . | 71      | 47       | 17         |
| Opinion . . . . .    | 27      | 6        | —          |

På ny kommer det klart frem at flertallet av våre fpp. har gitt sine svar på eksplisitt nivå, og at det er et svært lite antall som har gjort konsistent bruk av ett og samme nivå.

Det forhold at ingen har gjort konsistent bruk av opinion og implisitt nivå og meget få endog har overvekt på disse nivåer, begrenser sterkt våre muligheter for en systematisk analyse.

En løsning på problemet ligger i bruken av en mer relativ klassifiseringsmetode, hvor hovedvekten blir lagt på klassifisering av ekstrem-

grupper — i forhold til bruk av henholdsvis implisitt, eksplisitt og opinionsnivå.

Vår fremgangsmåte var følgende:

Som ekstremgruppe på eksplisitt nivå ble valt ut samtlige som hadde svart at de alltid gjorde bruk av eksplisitt nivå, dvs. hadde avmerket «Alltid» på spørsmål B. Som ekstremgruppe på opinionsnivå ble valt dem som hadde avmerket «Ofte» eller «Alltid» på spørsmål C, og som ekstremgruppe på implisitt nivå, dem som hadde avmerket «Iblant», «Sjelden» eller «Aldri» på spørsmålet A. Ved konflikt i klassifiseringen mellom avmerking av «Iblant» på spørsmål A og «Ofte» på spørsmål C ble opinionsnivå foretrukket, dvs. fp. ble utelukkende klassifisert på dette nivå.

Fremgangsmåten ga mulighet for utskilling av tre ekstremgrupper, henholdsvis med 16, 17 og 18 fpp. Tilsammen omfatter ekstremgruppene altså 51 eller 65 % av vårt totalutvalg på 79.

For utvalget som helhet viste det seg, som nevnt, utelukkende å være ego-defens og behovspersistens som ga noelunde statistisk pålitelige korrelasjoner med tilpassingsvurderingene. I den videre analyse har vi derfor bare konsentrert oss om disse to kategorier.

Tabell 19 gir en oversikt over korrelasjonen mellom testresultater og tilpassingsvurderinger beregnet særskilt for de tre ekstremgrupper.

TABELL 19

*Korrelasjonen (r) mellom 'tilpassing' og henholdsvis E-D og B-P i forskjellige ekstremgrupper*

| Kategori | Totalt utvalg<br>N=79 | Ekstrem opinion<br>N=17 | Ekstr. eksplisitt<br>N=16 | Ekstrem implisitt<br>N=18 |
|----------|-----------------------|-------------------------|---------------------------|---------------------------|
| E-D      | — .19                 | + .18                   | — .52                     | — .24                     |
| B-P      | + .20                 | + .14                   | + .21                     | + .18                     |

Tabellen viser at korrelasjonen varierer betydelig i de tre grupper. Høyest er korrelasjonen i de tilfellene B-F-testen er blitt besvart på eksplisitt nivå. Både E-D- og B-P-kategoriene viser her høyere korrelasjon med «tilpassing» enn i totalutvalget. Påfallende høyere er korrelasjonen med E-D-kategorien.

I ekstremgruppen med opinionsnivå er forskjellene fra totalutvalget større enn i ekstremgruppen med implisitt svarnivå. Særlig i forbindelse med E-D-kategorien er forskjellen påfallende. Mens det i alle de andre utvalgene er en negativ korrelasjon mellom ego-defensive tendenser og «tilpassing», er forholdet her det motsatte.

Statistisk mest pålitelig er forskjellene mellom ekstremgruppene med forankring på opinion og eksplisitt nivå. I forbindelse med E-D-kategorien får vi her forskjeller som er signifikant på under 5 %-nivået.



Rent skjematisk kan vi si at ekstremgruppen med eksplisitt svarnivå på testen viser størst korrelasjon, og ekstremgruppen med opinionsnivå minst korrelasjon ved tilpassing. Dette kommer også frem på tabell 20, som gir en oversikt over korrelasjonen mellom testresultater og tilpassingsvurderinger beregnet særskilt for underutvalgene som fremkom ved bruk av vår absolutte klassifiseringsmetode.

TABELL 20

*Korrelasjonen (r) mellom 'tilpassing' og E-D og B-P i forskjellige underutvalg*

| Nivå       | Kategori | Tendens | Overvekt | Konsistens |
|------------|----------|---------|----------|------------|
| Implisitt  | E-D      | — .19   |          |            |
|            | B-P      | + .22   |          |            |
| Eksplisitt | E-D      | — .20   | — .27    | — .45      |
|            | B-P      | + .16   | + .12    | + .16      |
| Opinion    | E-D      | + .02   | + .40    |            |
|            | B-P      | + .06   | — .38    |            |

Tabellen viser at E-D-kategorien korrelerer høyest med «tilpassing» i utvalget med eksplisitt svarnivå, og høyere jo mer konsistent dette nivå er blitt nyttet.

Når det gjelder konsistent bruk av implisitt og opinionsnivå, er våre data mangelfulle. Det er allikevel av interesse å konstatere at overvekt på eksplisitt nivå gir tilnærmet signifikant bedre korrelasjon enn overvekt på opinionsnivå ( $p = .10$ ). Likeledes at E-D-korrelasjonen på implisitt og opinionsnivå ikke avviker signifikant fra null, mens dette er tilfelle på eksplisitt nivå. Eksplisitt tendens gir  $p = .09$  og eksplisitt overvekt  $p = .05$ .

Påfallende er også at i utvalget med overvekt på opinionsnivå går korrelasjonene i motsatt retning av de tilsvarende korrelasjoner i de andre utvalg. Dette medfører selvsagt en sterk reduksjon av totalutvalgets korrelasjoner.

Særlig ufullstendig er vårt materiale i forbindelse med bruk av implisitt svarnivå. Bare en av våre fpp. har overvekt på dette nivå. Noen slutninger med hensyn til korrelasjonen mellom «tilpassing» og implisitte tendenser er det derfor vanskelig å trekke.

Vi antok innledningsvis at tilpassingsvurderingene hadde en eksplisitt nivåforandring, og at høyest korrelasjon mellom tilpassingsvurderinger og testresultater således skulle være til stede i de tilfelle testen ble besvart konsistent på eksplisitt nivå. Dette viser seg å være tilfelle.

Likeledes viser det seg at testens skikkethet til å forutsi tilpassingsvurderinger er avhengig av hvilket svarnivå som blir anvendt. Ved å ta hensyn til svarnivå synes testens validitet å bli betydelig økt. Ved bruk av eksplisitt nivå på testen får vi således korrelasjonskoeffisienter med kriteriet for E-D-kategorien på — .52 og — .45, alt etter hvilken metode vi nytter til nivå-klassifisering.

Vi har ovenfor antatt at vår metode til nivåbestemmelse er valid. Også tilpassingsvurderingenes eksplisitte nivåforankring bygger på antagelser. Vi har tidligere berørt at de muligens er lite reliable — sett på bakgrunn av at de er blitt foretatt av psykologer med forskjellig trening og erfaring i atfredsbedømmelse.

Alle disse forhold må selvsagt tas i betraktning ved en nærmere vurdering av våre resultater. Til tross for en rekke usikre momenter er det imidlertid vår oppfatning at vår validitetsundersøkelse indirekte gir et positivt bidrag til testens anvendbarhet.

## UNDERSØKELSE OVER KORRELASJONEN MED EN INNSTILLINGSSKALA

Som nevnt ble B-F-testen i vår undersøkelse brukt sammen med en «multiple choice» innstillingsskala. Dette ble gjort for å få et mest mulig allsidig grunnlag for bedømmelse av våre fpp.s måte å reagere på overfor interpersonlig frustrasjon. Vi håpet bl.a. å få en viss indre gjensidig kontroll på de to testers validitet. Dette er selvsagt en høyst utilfredsstillende fremgangsmåte, men vi synes at den allikevel var å foretrekke fremfor å arbeide uten enhver kontroll.

Det er tidligere blitt gitt en kort beskrivelse av innstillingsskalaen (kfr. s. 83). I ytre form synes den å ha visse likheter med del R i Rosenzweigs opprinnelige F-test (kfr. s. 71). Skåringskategoriene som blir anvendt, er imidlertid noe forskjellig.

Når det gjelder utviklingen av innstillingsskalaen, gikk vi frem på følgende måte: Først ble det konstruert et «spørreskjema» bestående av 60 ledd. Hvert ledd var en kort beskrivelse av en frustrerende situasjon som ble fulgt av en oppfordring til fp. om å angi hvorledes han eller hun trolig ville reagerer på situasjonen. Spørreskjemaet hadde m.a.o ingen fastlåste svaralternativer. Det ble administrert på et mindre utvalg studenter og gymnasiaster ( $N = 38$ ) våren 1952. På grunnlag av svarmaterialet ble det utprøvd forskjellige skåringsmetoder med sikte på å finne frem til et passende klassifiseringssystem. Til slutt konsentrerte vi oss som nevnt om en noe forenklet utgave av Rosenzweigs skåringskategorier, samtidig som vi besluttet i den endelige innstillingsskala utelukkende å gjøre bruk av 40 situasjoner, fortrinnsvis de som viste størst spredning med hensyn til svartyper. Vår neste oppgave var å trekke frem eksempler på de forskjellige reaksjonstyper på de enkelte situasjoner, samt velge ut svaralternativer til bruk i den endelige skala.

I et mindre antall tilfelle viste det seg nødvendig å konstruere svar.

Dette var i forbindelse med de situasjoner hvor det ved forprøven ikke forekom spontane svar i alle skåringskategorier.

Etter at svaralternativer var utvalgt eller utarbeidet, ble de skåret uavhengig av to personer. Hvis skåringen var ens og for øvrig i overensstemmelse med den kategori svaret var ment å dekke, ble svaralternativet beholdt; hvis ikke, ble det byttet ut med et nytt svar og skåringen gjort om igjen. Først når det var full enighet om skåringen av samtlige svaralternativer på alle ledd, ble innstillingsskalaen betraktet som endelig.

Instruksjonen til innstillingsskalaen sier bl.a.:

«I dette spørreskjema vil du bli stilt overfor i alt 40 forskjellige situasjoner. Etter hver situasjon er det satt opp 6 forskjellige reaksjoner. Noen av disse blir selvsagt brukt hyppigere enn andre, men de er alle vanlige reaksjoner. Du vil bli anmodet om å angi på hvilke av de nevnte måter du trolig vil reagere.

Du skal gi to svar i forhold til hver situasjon: på hvilken av de nevnte måter du anser det aller mest trolig, og dernest mest trolig at du vil reagere. Svarene dine skal du ikke skrive i dette skjemaet, men på et særskilt svarark. Du skal sette en sirkel om de bokstavene på svararket som står for de av svaralternativene etter hver enkelt situasjon som du synes best dekker dine trolige reaksjoner. Her er gitt et eksempel:

100 mest trolig            a b (c) d e f  
                                  dernest trolig        a b c (d) e f

I dette tilfelle (situasjon 100) ville svaralternativ «c» være den reaksjon blant de nevnte som du anser aller mest trolig, og «d» den du anser dernest mest trolig.

Det er ingen tidsbegrensning, men det er av betydning at du arbeider hurtig. Noen ganger vil du sikkert føle deg usikker på hva du skal svare. Stopp da aldri opp, men velg det alternativ som først faller deg inn som mest sannsynlig. Unngå enhver inngående gransking av dine egne reaksjoner og forsøk å arbeide mest mulig sammenhengende. Svar ikke hva du bør gjøre, men hva du faktisk vil gjøre i situasjonen.»

Nedenfor er et par eksempler hentet direkte fra skalaen. I parentes er angitt hvordan de forskjellige svaralternativer var blitt preskåret. Dette var selvsagt ikke påført i de prøvehefter som fpp. gjorde bruk av.

13. Du har bestilt bord og invitert en venn med på en danserestaurant. Da dere ankommer, får dere vite av hovmesteren at han nok har glemt bestillingen og at det ikke er noe bord ledig.

På hvilken av følgende måter vil du trolig reagere?

- a) Føle meg bedrøvet over å ha satt hovmesteren i en slik knipe, bestille bord en annen dag. (i)
- b) Vente og se om et bord skulle bli ledig. (m)
- c) Bli irritert på meg selv for å ha bestilt bord på en slik restaurant. (I)
- d) Henvende meg til direktøren, forlange at det blir skaffet et nytt bord. (e)

- e) Unnskyldte hovmesteren, det er menneskelig å glemme. (M)
- f) Bli sint og irettesette hovmesteren for hans glemsomhet. (E)

21. En venn av deg legger ubetenksomt en glødende sigarett på bordet ditt og brenner et stort hull både i bordduken og i selve bordplaten.

På hvilken av følgende måter vil du trolig reagere?

- a) Bebreide vedkommende, be ham se seg bedre for en annen gang. (E)
- b) Be vedkommende erstatte skaden eller sørge for at den blir reparert. (e)
- c) Bli ergerlig på meg selv at jeg ikke hadde vært mer påpasselig. (I)
- d) Bli lei meg på vennens vegne, tilby å fikse det selv. (i)
- e) Et hendelig uhell, slikt kan hende den beste. (M)
- f) Ta det hele med ro. (m)

Det vil føre for langt her å gi en fyldig beskrivelse av hvorledes skåresummering og tabulering ble utført. En slik beskrivelse er det meningen å gi senere i en annen publikasjon. I denne vil det bli gitt en fullstendig fremstilling av innstillingsskalaen, en nærmere omtale av hvorledes skalaen ble konstruert, og en beskrivelse av fremgangsmåtene som ble anvendt ved beregning av de forskjellige skåringskategoriers diskrimineringssevne, indre konsistens, etc.

Det er særlig én side ved innstillingsskalaen vi her vil trekke frem. Både i instruksjonen og ved oppfordring etter hver situasjonsbeskrivelse blir fpp. anmodet om å gi sine svar på eksplisitt atferdsnivå. Vi har ingen mulighet til direkte å kontrollere om dette faktisk blir etterfulgt. Vi er utelukkende henvist til å anta at så er tilfelle.

Det første spørsmål vi tok opp til analyse, var forholdet mellom innstillingsskalaen og B-F-testen. På grunn av B-F-testens nivåabilitet var vi innstilt på å finne en positiv, men begrenset korrelasjon.

TABELL 21

Korrelasjonen ( $r$ ) mellom B-F-testen og innstillingsskalaen ( $N = 135$ )

|     | Ex   | In   | Im   | E-D | B-P | Mdn  |
|-----|------|------|------|-----|-----|------|
| $r$ | .43  | .28  | .32  | .11 | .16 | .28  |
| $p$ | .001 | .001 | .001 | .12 | .07 | .001 |

Tabell 21 viser at dette i høy grad er tilfelle. Den mest bemerkelsesverdige side ved resultatene er den til dels betydelig høyere korrelasjon med retningskategoriene enn med formkategoriene. De første viser korrelasjoner som alle er signifikante på under .001-nivået, mens de siste bare oppnår signifikans på omkring 10%-nivået.

En mulig forklaring på korrelasjonsforskjellene er at formkategoriene stort sett er mindre reliable enn retningskategoriene.

For å få kontroll på dette har vi foretatt en særlig undersøkelse.

Ved å regne test-kategoriernes reliabilitet (indre konsistens) er det mulig å korrigere korrelasjonene mellom dem slik at vi får uttrykk for den korrelasjon som er til stede hvis kategoriene hadde vært uten målingsfeil (correction for attenuation). Reliabiliteten til de forskjellige kategorier har vi beregnet ved hjelp av Hoyts variansanalytiske metode.

TABELL 22

Oversikt over innstillingsskalaens og B-F-testens indre konsistens ( $r_{tt}$ )

| Metode              | Antall ledd | N   | Ex  | In  | Im  | H-D | E-D | B-P |
|---------------------|-------------|-----|-----|-----|-----|-----|-----|-----|
| B-F-testen . . . .  | 24          | 136 | .53 | .14 | .40 | .28 | .27 | .43 |
| Innstillingsskalaen | 40          | 165 | .89 | .71 | .78 | —   | .68 | .68 |

Når vi korjierer for de forskjellige kategoriernes mangelfulle reliabilitet, fremkommer de korrelasjoner som er angitt i tabell 23.

TABELL 23

Korrelasjonen mellom «sanne» variabler ( $N = 135$ )

|            | Ex  | In  | Im  | E-D | B-P | $Mdn$ |
|------------|-----|-----|-----|-----|-----|-------|
| $r_{korr}$ | .63 | .92 | .59 | .25 | .30 | .59   |

Tabellen viser at det fremdeles er påfallende forskjeller mellom retnings- og formkategoriene.

Fra vår tidligere undersøkelse over forholdet mellom B-F-testen og tilpassingsvurderinger har vi data som antyder at i det minste formkategoriene er høyst nivålabile. En mulig forklaring på korrelasjonsforskjellene ovenfor er derfor at formkategoriene er mer nivålabile enn retningskategoriene.

Denne antagelse kan vi også underkaste empirisk prøving.

Hvis innstillingsskalaen registrerer eksplisitte atferdstendenser, skulle vi vente en høyere korrelasjon mellom innstillingsskalaen og B-F-testen, jo mer vi konsentrerer oss om utvalg som har besvart B-F-testen på eksplisitt nivå. For utvalgene som har gjort bruk av implisitt eller opinionsnivå på B-F-testen, skulle vi vente lave korrelasjoner, og relativt lavere for formkategoriene enn for retningskategoriene.

Til klassifisering av våre fpp.s svarnivå på B-F-testen har vi gjort bruk av den relative fremgangsmåte som vi tidligere har beskrevet, dvs. vi har tatt for oss ekstremgrupper på henholdsvis implisitt, eksplisitt og opinionsnivå.

Ved bruk av de samme kriterier får vi nå på grunn av det større totalutvalg tre ekstremgrupper på henholdsvis 33, 24 og 23 forsøks-  
personer. Tilsammen utgjør ekstremgruppene 59 % av totalutvalget.

Tabell 24 gir en oversikt over korrelasjonen mellom innstillings-  
skalaen og B-F-testen beregnet særskilt for de tre ekstremgrupper.

TABELL 24

*Korrelasjonen (r) mellom innstillingsskalaen og B-F-testen  
i forskjellige ekstremgrupper*

| Kategori   | Total utvalg<br>(N=135) | Ekstr. opinion<br>(N=23) | Ekstr. eksplisitt<br>(N=24) | Ekstr. implisitt<br>(N=33) |
|------------|-------------------------|--------------------------|-----------------------------|----------------------------|
| Ex         | .43                     | .41                      | .68                         | .52                        |
| In         | .28                     | .38                      | .42                         | .48                        |
| Im         | .32                     | .48                      | .54                         | .30                        |
| E-D        | .11                     | .10                      | .48                         | .14                        |
| B-P        | .16                     | -.15                     | .34                         | .12                        |
| <i>Mdn</i> | .28                     | .38                      | .48                         | .30                        |

Tabellen viser at ekstremgruppen med bruk av eksplisitt svarnivå på B-F-testen oppnår høyere korrelasjoner enn alle andre utvalg i forbindelse med alle kategorier, på én unntagelse nær. Når det gjelder In-kategorien, ligger ekstremgruppen med implisitte svar noe høyere.

Tabellen antyder for øvrig på ny betydningen av å ha kjennskap til hvilket svarnivå som anvendes på B-F-testen. Av særlig betydning synes et slikt kjennskap å være når det gjelder vurdering av formkategoriene. Mens korrelasjonene i forbindelse med disse kategorier i utvalget med eksplisitt svarnivå ikke avviker stort fra retningskategorienes korrelasjoner, er dette i høy grad tilfelle i de andre utvalgene.

Vår antagelse med hensyn til formkategorienes relativt større nivå-labilitet blir altså støttet.

For å få en ytterligere kontroll på holdbarheten av denne konklusjon har vi også beregnet korrigerede korrelasjoner for de tre ekstremgrupper. Tabell 25 gir en oversikt over våre resultater.

Tabellen antyder at formkategoriene er mer nivå-labile enn retningskategoriene. Utelukkende i ekstremgruppen med eksplisitt svarnivå er formkategorienes korrelasjoner ikke utpreget lavere enn retningskategorienes. Medianverdien i denne gruppe er .99, noe som for øvrig antyder at innstillingsskalaen og B-F-testen stort sett måler nøyakig

de samme egenskaper — i de tilfelle testen blir besvart på eksplisitt nivå.

TABELL 25

*Korrigert korrelasjon mellom innstillingsskalaen og B-F-testen i forskjellige ekstremgrupper*

| Kategori   | Total utvalg<br>(N=135) | Ekstr. opinion<br>(N=23) | Ekstr. eksplisitt<br>(N=24) | Ekstr. implisitt<br>(N=33) |
|------------|-------------------------|--------------------------|-----------------------------|----------------------------|
| Ex         | .63                     | .60                      | .99                         | .76                        |
| In         | .92                     | (1.00)                   | (1.00)                      | (1.00)                     |
| Im         | .59                     | .88                      | .99                         | .55                        |
| E-D        | .25                     | .19                      | .96                         | .28                        |
| B-P        | .30                     | -.29                     | .64                         | .23                        |
| <i>Mdn</i> | .59                     | .60                      | .99                         | .55                        |

Hadde de to ekstremgruppene med hensyn til bruk av opinion og implisitt nivå vært mer ekstreme, dvs. hadde begge bare omfattet personer som hadde gjort konsistent bruk av enten opinion eller implisitt svarnivå, kunne våre data gitt holdepunkter til vurdering av forholdet mellom reaksjonstendenser på henholdsvis eksplisitt og implisitt, og eksplisitt og opinions-nivå. Som nevnt er det ytterst få personer i vårt totalutvalg som har gjort konsistent bruk av disse nivåer. Noen beregning av korrelasjonen mellom atferdsnivåer gir vårt materiale derfor ikke grunnlag for.

En forutsetning for beregningen av de korrigerede korrelasjonene ovenfor er at kategori-reliabiliteten for de forskjellige utvalg er den samme som for totalutvalget. Vi har nemlig konsekvent gjort bruk av de reliabilitetsverdier som ble beregnet for utvalget som helhet.

Når det gjelder innstillingsskalaen, er denne forutsetning trolig holdbar, mens den i forbindelse med B-F-testen er mer usikker.

I alle ekstremgrupper oppnår den korrigerede korrelasjon for In-kategorien høyere verdier enn 1.00. Dette antyder at kategori-reliabiliteten i totalutvalget er lavere enn i underutvalgene. I særlig grad synes dette å være tilfelle for ekstremgruppen med eksplisitt svarnivå.

Når vi særlig fester oss ved B-F-testen i denne forbindelse, er det på grunn av dens relativt lave reliabilitet sett i forhold til dens sterke nivåabilitet.

Det er ikke usannsynlig at B-F-testens lave reliabilitet nettopp gir uttrykk for at en meget stor del av våre fpp. har gitt inkonsistente svar, dvs. svar på forskjellige reaksjonsnivåer. Det er således tenkelig at

reliabiliteten hadde vært betydelig høyere hadde testens svarnivå vært mer fiksert. Fra andre undersøkelser vet vi at korrelasjonen mellom reaksjoner på eksplisitt og opinions-nivå ofte er lav, og likeledes at forholdet mellom implisitt og eksplisitt nivå dels er blitt påvist å være negativt korrelert, dels ukorrelert (kfr. s. 74 f.).

En sammenligning med innstillingsskalaens reliabilitet støtter vår hypotese. Ved å korrigere for B-F-testens mindre antall ledd får vi et reelt grunnlag for en slik sammenligning. Tabell 26 gir en oversikt over forholdet mellom de to metoders reliabilitet.

TABELL 26

*Sammenligning mellom B-F-testens og innstillingsskalaens reliabilitet*

| Metode             | Antall ledd | N   | Ex  | In  | Im  | H-D | E-D | B-P |
|--------------------|-------------|-----|-----|-----|-----|-----|-----|-----|
| B-F-testen . . . . | 40          | 136 | .65 | .21 | .52 | .39 | .38 | .56 |
| Innst. skalaen ..  | 40          | 165 | .89 | .71 | .78 | —   | .68 | .68 |

Både ved B-F-testen og innstillingsskalaen er det den ekstrapunitive kategori som oppnår høyest reliabilitet. Ekstrapunitivitet synes mer enn noen av de andre kategoriene å stå for et relativt konsistent personlighetstrekk: Individuer som, sammenlignet med andre, er relativt mer aggressive i én situasjon, synes også å være relativt mer aggressive i andre situasjoner. Også i en nyere amerikansk undersøkelse er det blitt påpekt at aggressive holdninger er betydelig mindre situasjons-spesifikke enn ofte antatt (26).

En systematisk sammenligning av B-F-testens og innstillingsskalaens reliabilitet viser at innstillingsskalaen ligger høyere på alle kategorier.

Hypotesen om at B-F-testens lave reliabilitet skyldes dens nivårelabilitet kan også undersøkes mer direkte. Vi kan f.eks. ta for oss det underutvalg som har besvart testen på mest inkonsistent grunnlag, og beregne reliabilitetsverdier spesielt for dette utvalg. Vi skulle i så fall her vente å finne en lavere reliabilitet enn i totalutvalget.

Optimal inkonsistens i bruk av svarnivå foreligger i de tilfelle de forskjellige svarnivåer er blitt nyttet like hyppig, eller mer operasjonelt definert — i de tilfelle fp. er blitt tillagt den samme «ladning» på alle tre nivåer ( $A = B = C$ ).

I alt omfatter dette underutvalg 20 fpp. Tabell 27 gir en oversikt over de forskjellige kategoriens reliabilitet beregnet spesielt for dette utvalg. Også her er anvendt Hoyts metode.

Til sammenligning er også anført reliabilitetsverdiene for totalutvalget. For 4 av testens 6 kategorier ligger totalutvalget høyere. Både i forbindelse med In- og B-P-kategorien er underutvalgets reliabilitetskoeffisient praktisk talt null.



TABELL 27

*Sammenligning mellom 'inkonsistens'-utvalgets og totalutvalgets reliabilitet*

| Utvalg                 | N   | Ex  | In   | Im  | H-D | E-D | B-P  | M   |
|------------------------|-----|-----|------|-----|-----|-----|------|-----|
| Optimal inkonsistens.. | 20  | .24 | -.04 | .21 | .60 | .44 | -.03 | .24 |
| Total .....            | 136 | .53 | .14  | .40 | .28 | .27 | .43  | .34 |

Det er av stor betydning å være oppmerksom på at underutvalget bare omfatter 20 personer. Resultatene er derfor høyst upålitelige. De støtter stort sett vår hypotese, men gir ikke grunnlag for noen definitiv konklusjon.

En annen fremgangsmåte som kan anvendes, er å konsentrere seg om det utvalg som har gjort bruk av et konsistent eksplisitt svarnivå. I dette tilfelle skulle vi vente en viss overensstemmelse med innstillingskalaens reliabilitetsverdier. Feilkildene på grunn av et lite antall fpp. vil imidlertid fremdeles være til stede, endog i enda større grad enn i første tilfelle. Vårt totalutvalg er i seg selv for lite til at vi kan vente særlig pålitelige resultater av undersøkelser av den art vi her står overfor.

Hvis vi antar holdbarheten av at B-F-testens lave reliabilitet til en viss grad skyldes dens høye nivåabilitet, skulle vi altså vente en mer tilfredsstillende reliabilitet i den grad svar på forskjellige nivåer blir særskilt behandlet. Av dette følger at det istedenfor 6 skåringskategorier, 3 reaksjonsretninger og 3 reaksjonsformer bør innføres hele 18 kategorier, 6 kategorier for henholdsvis implisitt, eksplisitt og opinionsnivå. Til hvert enkelt av de foreliggende skåringstegn er det ønskelig å tilknytte en særlig indeks som i tillegg angir hvilket svarnivå som er blitt nyttet.

Noen spesiell metode er ennå ikke blitt utarbeidet som kan gjøre en slik dobbeltskåring mulig. En vei å gå vil være å administrere forskjellige deler av testen med forskjellig instruksjon. Dette er imidlertid heller ikke noen helt sikker fremgangsmåte, og behovet for å følge opp med mer direkte spørsmål om svarnivå vil sannsynligvis fortsatt være til stede.<sup>25</sup>

Den framgangsmåte som er blitt anvendt i den foreliggende undersøkelse, gir sannsynligvis svært grove indikasjoner. Til bruk i forsk-

<sup>25</sup> Det åpner seg her høyst interessante forskningsoppgaver. Hvis det skulle vise seg mulig å bestemme svarnivå gjennom forandringer av instruksjonen, ville det være ønskelig å standardisere testen på forskjellige nivåer. Testens praktiske anvendbarhet ville bli betydelig økt. Det ville f.eks. kunne føre til at en kunne tilfredsstille en personalsjef som sa han hovedsaklig var interessert i fp.s eksplisitte reaksjonsnivå, og også klinikerens — som i første rekke var interessert i implisitte reaksjonstendenser. Strengt tatt er det ikke alltid behov for å prøve en fp. på alle nivåer.

ningsmessig øyemed i arbeid med grupper synes den å være et hensiktsmessig instrument, mens den derimot ved individuell prøving i en klinisk sammenheng sannsynligvis ofte vil komme til kort. I forbindelse med individuelle undersøkelser vil vi riktignok ha den fordel at vi gjennom hele testen vil kunne observere fp., vi vil ofte kunne hente informasjoner fra andre kilder, vi vil kunne knytte testingen sammen med et personlig intervju, vi vil kunne spørre fp. direkte eller indirekte, ledd for ledd, hvilket svarnivå er blitt nyttet, og vi vil kunne utnytte en hel rekke åpne og skjulte «tegn» (f.eks. reaksjonstid) som det er overordentlig vanskelig å få tak i i en gruppesituasjon.

Tallrike og forskjelligartede «tegn» vil kunne gi oss gode antydninger om svarnivå. Det er imidlertid én side ved denne framgangsmåten vi bør merke oss. Framgangsmåten vil i sin form ikke være psykometrisk; den vil ikke representere noen tallmessig registrerbar metode. Den vil i første rekke stille krav til testadministratorens psykologiske vurderingsevne og innsikt. Den vil i høy grad være kvalitativ. Fra dette perspektiv kan man med en viss rett hevde at det utelukkende er den erfarne kliniske psykologs bruk av testen som kan gi grunnlag for adekvate validitets- og reliabilitetsundersøkelser.

## SAMMENDRAG OG KONKLUSJONER

Med sikte på å klargjøre B-F-testens vitenskapelige verdi har vi foretatt en rekke spesielle empiriske undersøkelser. Undersøkelsene gjelder 1) testens skåringsreliabilitet, 2) testens indre konsistens, 3) forholdet mellom våre testnormer og tilsvarende amerikanske, 4) testens differensieringsevne mellom vårt utvalg og et utvalg skolehjemselever, 5) testens korrelasjon med tilpassingsvurderinger basert på intervjuopplysninger, 6) testens nivåabilitet, 7) testens korrelasjon med en spesiell innstillingsskala, og 8) testens validitet under hensyntagen til dens nivåabilitet. Våre resultater gir grunnlag for følgende konklusjoner:

1. Frustrasjonstesten synes stort sett å ha en tilfredsstillende skåringsreliabilitet. Ledd nr. 8 og 17 byr imidlertid relativt ofte på skåringsvansker. En utskiftning av disse ledd vil kunne føre til at testens skåringsreliabilitet blir bedret. Når det gjelder bruken av de amerikanske skåringseksemplene, har vi kommet til at disse med fordel kan anvendes i norsk oversettelse.

2. Alle skåringskategorier unntatt In-kategorien synes å være i stand til å differensiere mellom personer. Testen gir imidlertid ikke særlig pålitelige resultater. Fra en psykometrisk synsvinkel er skåringskategoriene lite homogene og det grunnlag testen gir til vurdering av enkeltindivider, relativt dårlig. Til en viss grad synes den meget begrensede indre konsistens som preger de fleste av kategoriene, å kunne tilskrives en påfallende liten differensieringsevne ved enkelte ledd. En utskiftning av ledd nr. 7, 10 og 19 og en generell økning av antall ledd vil sannsynligvis kunne bringe positive resultater.

3. Sammenligning med forskjellige amerikanske testnormer gir grunn til å anta at testsituasjonen er av betydning, og at Rosenzweigs normmateriale har betydelig begrensninger. I mangel av særlige norske testnormer synes de amerikanske å kunne gi visse generelle holdepunkter under forutsetning av at man velger normer med omtanke og forsøker å se det hele i en større sammenheng.

4. Testen synes å ha god differensieringsevne overfor forskjellige sosiale grupper. Ved sammenligning mellom vårt utvalg og et utvalg skolehjemsgutter påvises en rekke signifikante forskjeller som alle er i overensstemmelse med hva vi skulle vente å finne ut fra teoretiske betraktninger.

5. Testen viser liten korrelasjon med tilpassingsvurderinger foretatt av uttakspsykologer ved Forsvarets Psykologiske Avdeling. Størst er korrelasjonen med E-D- og B-P-kategorien på testen. Den multiple korrelasjon mellom disse to kategorier oppnår en viss størrelse, men korrelasjonen er allikevel så lav at forutsigelser basert på B-F-testen alene vil være usikre.

6. Testen synes å være et meget nivålabilt instrument. Til tross for at vårt utvalg med fpp. var relativt ensartet og testsituasjonen stort sett var den samme, antyder våre data at det foreligger en god del variasjon i fpp.s svarnivå på testen. I de fleste tilfelle er det sannsynligvis blitt anvendt et eksplisitt svarnivå, men det synes også å være dem som hovedsakelig har vært preget av andre nivåer. Det synes videre å foreligge en ikke ubetydelig intraindividuell variasjon. Behovet for metoder til nivåbestemmelse som gjør det mulig å spesifisere denne nærmere, er i høy grad til stede.

7. Testen viser en positiv, men relativt lav korrelasjon med en verbal innstillingsskala. Det foreligger imidlertid påfallende forskjeller mellom form- og retningskategoriene, idet retningskategoriernes korrelasjoner alle oppnår en betydelig statistisk pålitelighet.

8. Testens validitet synes i meget høy grad å være avhengig av at det blir tatt hensyn til hvilket svarnivå som blir anvendt i det enkelte tilfelle. Både korrelasjonen med tilpassingsvurderinger og innstillingsskalaen viser seg å bli betydelig økt resp. redusert når det blir foretatt særskilte korrelasjonsberegninger for undervalg som har besvart testen på forskjellige nivåer. For øvrig antyder våre data at formkategoriene på testen er mye mer nivålabile enn retningskategoriene. Testens utilfredsstillende indre konsistens synes til en viss grad å kunne tilskrives testens relativt store nivålabilitet.

## Litteratur

1. **Abt, L. E., & Bellak, L.** Projective psychology. New York: Knopf, 1952.
2. **Albee, G. W.** Patterns of aggression in psychopathology. *J. consult. Psychol.*, 1950, **14**, 465—468.
3. **Albee, G. W., & Goldman, R.** The Picture Frustration Study as a predictor of overt aggression. *J. proj. Tech.*, 1950, **14**, 303—308.
4. **Barker, R. G., Dembo, T., & Lewin, K.** Frustration and regression: An experiment with young children. *Univer. Iowa Stud. Child Welfare*, 1941, **18**, 1.
5. **Bell, J. E.** The case of Gregor: Psychological test data. *Rorschach Res. Exch. & J. proj. Tech.*, 1949, **13**, 155—205.
6. **Bernard, J.** The Rosenzweig Picture-Frustration Study: I. Norms, reliability, and statistical evaluation. *J. Psychol.*, 1949, **28**, 325—332.
7. **Bernard, J.** The Rosenzweig Picture-Frustration Study: II. Interpretation. *J. Pers.*, 1949, **28**, 333—343.
8. **Brown, J. F.** A modification of the Rosenzweig Picture-Frustration Test to study hostile interracial attitudes. *J. Psychol.*, 1947, **24**, 247—272.
9. **Brown, R. L., & Lacey, O. L.** The diagnostic value of the Rosenzweig P-F Study. *J. clin. Psychol.*, 1954, **10**, 72—75.
10. **Brun-Gulbrandsen, S.** Bidrag til utforskningen av de psykologiske problemer som knytter seg til skolehjemselevne. Magistergradsavhandling, Oslo, 1951.
11. **Buros, O. K.** (Ed.) The fourth mental measurements yearbook. Highland Park, N. J.: Gryphon Press, 1953.
12. **Challman, R. C.** Rosenzweig Picture-Frustration Study: A review. *I Buros* (11), pp. 240—242.
13. **Clarke, H. J.** The diagnosis of a patient with limited capacity. *J. Pers.*, 1946, **15**, 105—112.
14. **Clarke, H. J., Rosenzweig, S., & Flemming, E. E.** The reliability of the scoring of the Rosenzweig Picture-Frustration Study. *J. clin. Psychol.*, 1947, **3**, 364—370.
15. **Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., Sears, R. R., Ford, C. S., Hovland, C. I., & Sollenberger, R. T.** Frustration and aggression. New Haven: Yale Univer. Press, 1939.

16. **Dressel, P. L.** Some remarks on the Kuder-Richardson reliability coefficient. *Psychometrika*, 1940, **5**, 305—310.
17. **Du Bois, P. H., & Watson, R. I.** The selection of patrolmen. *J. appl. Psychol.*, 1950, **34**, 90—95.
18. **Falls, R. P., & Blake, R. R.** A quantitative analysis of the Picture-Frustration Study. *J. Pers.*, 1948, **16**, 320—325.
19. **Farberow, N. L.** Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, **42**, 3—79.
20. **Fisher, S., & Hinds, E.** The organization of hostility controls in various personality structures. *Genet. Psychol. Monogr.*, 1951, **44**, 3—68.
21. **Franklin, J. C., & Brozek, J.** The Rosenzweig P-F Test as a measure of frustration responses in semistarvation. *J. consult. Psychol.*, 1949, **13**, 293—301.
22. **French, R. L.** Changes in performance on the Rosenzweig Picture-Frustration Study following experimentally induced frustration. *J. consult. Psychol.*, 1950, **14**, 111—115.
23. **Fry, D.** A study of reaction to frustration in 236 college students and in 207 inmates of state prisons. *J. Psychol.*, 1949, **28**, 427—438.
24. **Gatling, F. P.** Frustration reactions of delinquents using Rosenzweigs classification system. *J. abnorm. soc. Psychol.*, 1950, **45**, 749—752.
25. **Goodenough, F. L.** *Mental Testing*. New York: Rinehart, 1950.
26. **Graham, F. K., Charwat, W. A., Honig, A. S., & Wertz, P. C.** Aggression as a function of the attack and the attacker. *J. abnorm. soc. Psychol.*, 1951, **46**, 512—520.
27. **Gulliksen, H.** The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, **10**, 79—91.
28. **Harrison, R., & Rotter, J. B.** A note on the reliability of the Thematic Apperception Test. *J. abnorm. soc. Psychol.*, 1945, **40**, 97—99.
29. **Holzberg, J. D., Cahen, E. K., & Wilk, E. K.** Suicide: A psychological study of self-destruction. *J. proj. Tech.*, 1951, **15**, 339—354.
30. **Holzberg, J. D., & Posner, R.** The relationship of extrapunitiveness on the Rosenzweig Picture-Frustration Study to aggression in overt behavior and fantasy. *Amer. J. Orthopsychiat.*, 1951, **21**, 767—779.
31. **Holzberg, J. D., & Hahn, F.** The Picture-Frustration technique as a measure of hostility and guilt reactions in adolescent psychopaths. *Amer. J. Orthopsychiat.*, 1952, **22**, 776—797.
32. **Horst, P.** Pattern analysis and configural scoring. *J. clin. Psychol.* 1954, **10**, 3—11.
33. **Hoyt, C.** Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153—160.
34. **Hybl, A. R., & Stagner, R.** Frustration tolerance in relation to diagnoses and therapy. *J. consult. Psychol.*, 1952, **16**, 163—170.
35. **Johnson, P. O.** *Statistical methods in research*. New York: Prentice-Hall, 1949.
36. **Karlin, L., & Schwartz, M.** Social and general intelligence and perfor-

- mance on the Rosenzweig Picture-Frustration Study. *J. consult. Psychol.*, 1953, **17**, 293—296.
37. **Kogan, K. L.** The diagnosis of a patient with organic defects. *J. Pers.*, 1946, **15**, 113—120.
  38. **Korner, A. F.** Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619—267.
  39. **Korner, A. F.** Relationship between overt and covert hostility: Economy and dynamics. *Personality*, 1951, **1**, 20—31.
  40. **Lindzey, G.** An experimental test of the validity of Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 315—320.
  41. **Lindzey, G., & Riecken, H. W.** Inducing frustration in adult subjects. *J. consult. Psychol.*, 1951, **15**, 18—23.
  42. **Lindzey, G., & Goldwyn, R. M.** Validity of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1954, **22**, 519—547.
  43. **Maier, N. R. F.** *Frustration. A study of behavior without a goal.* New York: McGraw-Hill, 1949.
  44. **Maslow, A. H.** Deprivation, threat and frustration. *Psychol. Rev.*, 1941, **48**, 364—366.
  45. **Maslow, A. H.** Conflict, frustration and the theory of threat. *J. abnorm. soc. Psychol.*, 1943, **38**, 81—86.
  46. **McCary, J. L.** Ethnic and cultural relations to frustration. *J. Pers.*, 1950, **18**, 321—326.
  47. **McCary, J. L.** Reacting to frustration by some cultural and racial groups. *Personality*, 1951, **1**, 84—102.
  48. **McNemar, Q.** *Psychological statistics.* New York: Wiley, 1949.
  49. **Murray, H. A.** *Thematic Apperception Test manual.* Cambridge: Harvard Univer. Press, 1943.
  50. **Mussen, P. H., & Naylor, H. K.** The relationship between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, **49**, 235—240.
  51. **Newcomb, T. M.** *Social psychology.* New York: Dryden Press, 1950.
  52. **Quay, H., & Sweetland, A.** The relationship of the Rosenzweig P-F Study to the MMPI. *J. clin. Psychol.*, 1954, **10**, 296—297.
  53. **Rosenzweig, S.** Preferences in the repetition of successful and unsuccessful activities as a function of age and personality. *J. genet. Psychol.*, 1933, **42**, 423—441.
  54. **Rosenzweig, S.** The recall of finished and unfinished tasks as affected by the purpose with which they were performed. *Psychol. Bull.*, 1933, **30**, 698.
  55. **Rosenzweig, S., & Mason, G.** An experimental study of memory in relation to the theory of regression. *Br. J. Psychol.*, 1934, **24**, 247—265.
  56. **Rosenzweig, S.** Types of reactions of frustration. An heuristic classification. *J. abnorm. soc. Psychol.*, 1934 **29**, 298—300.
  57. **Rosenzweig, S.** A test for types of reactions to frustration. *Amer. J. Orthopsychiat.*, 1935, **5**, 395—403.

58. **Rosenzweig, S.** The experimental study of repression. In *Exploration in Personality*, ed. H. A. Murray. New York: Oxford Univer. Press, 1938, pp. 472—490.
59. **Rosenzweig, S.** The experimental measurement of types of reaction to frustration. In *Explorations in Personality*, ed. H. A. Murray. New York: Oxford Univer. Press, 1938, pp. 585—599.
60. **Rosenzweig, S., Mowrer, O. H., Haslund, G. M., Curtis, Q. F., & Barker, R. G.** Frustration as an experimental problem. *Charact. & Pers.*, 1938, **7**, 126—170.
61. **Rosenzweig, S.** Need-persistent and ego-defensive reactions to frustration as demonstrated by an experiment on repression. *Psychol. Rev.*, 1941, **48**, 347—349.
62. **Rosenzweig, S.** Fantasy in personality and its study by test procedures. *J. abnorm. soc. Psychol.*, 1942, **37**, 40—51.
63. **Rosenzweig, S., & Sarason, S.** An experimental study of the triadic hypothesis: Reaction to frustration, ego-defense and hypnotizability. *Charact. & Pers.*, 1942, **11**, 1—19.
64. **Rosenzweig, S.** An outline of frustration theory. In *Personality and Behavior Disorders*, ed. McV. Hunt. New York: Ronald Press, 1944, vol. 1, pp. 379—388.
65. **Rosenzweig, S.** The picture-association method and its application in a study of reaction to frustration. *J. Pers.*, 1945, **14**, 3—23.
66. **Rosenzweig, S., & Clark, R. A.** The personality of a psychotic exsoldier. *J. abnorm. soc. Psychol.*, 1945, **40**, 195—204.
67. **Rosenzweig, S.** The dynamics of an amnesic personality. *J. Pers.*, 1946, **15**, 121—142.
68. **Rosenzweig, S., Clarke, H. J., Garfield, M.S., & Lehndorff, A.** Scoring samples for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1946, **21**, 45—72.
69. **Rosenzweig, S., Flemming, E. E., & Clarke, H. J.** Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, **24**, 165—208.
70. **Rosenzweig, S., Flemming, E. E., & Rosenzweig, L.** The children's form of the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1948, **26**, 141—191.
71. **Rosenzweig, S., & Kogan, K.** *Psychodiagnosis. An introduction to the integration of tests in dynamic clinical practice.* New York: Grune & Stratton, 1949.
72. **Rosenzweig, S.** Frustration tolerance and the Picture-Frustration Study. *Psychol. Serv. Center J.*, 1950, **2**, 109—115.
73. **Rosenzweig, S.** Levels of behavior in psychodiagnosis with special reference to the Picture-Frustration Study. *Amer. J. Orthopsychiat.*, 1950, **20**, 63—72.
74. **Rosenzweig, S.** Some problems relating to research on the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 303—305.
75. **Rosenzweig, S.** The treatment of humorous responses in the Rosen-

- zweig Picture-Frustration Study: A note on the revised (1950) instructions. *J. Psychol.*, 1950, **30**, 139—143.
76. **Rosenzweig, S.** A method of validation by successive clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 507—509.
  77. **Rosenzweig, S.** Revised norms for the adult form of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 344—346.
  78. **Rosenzweig, S., & Mirmow, E. L.** The validation of trends in the children's form of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, **18**, 306—314.
  79. **Rosenzweig, S.** Idiodynamics in personality theory with special reference to projective methods. *Psychol. Rev.*, 1951, **58**, 213—223.
  80. **Rosenzweig, S., & Rosenzweig, L.** Aggression in problem children and normals as evaluated by the Rosenzweig P-F Study. *J. abnorm. soc. Psychol.*, 1952, **47**, 683—687.
  81. **Sanford, F. H., & Rosenstock, I. M.** Projective techniques on the doorstep. *J. abnorm. soc. Psychol.*, 1952, **47**, 3—16.
  82. **Sanford, R. N., et al.** Physique, personality and scholarship. *Monogr. Soc. Res. Child Developm.*, 1943, **8**, no. 1
  83. **Sargent, S. S.** Reaction to frustration — A critique and hypothesis. *Psychol. Rev.*, 1948, **55**, 108—114.
  84. **Schmeidler, G. R.** Some relations between Picture-Frustration ratings and ESP scores. *J. Pers.*, 1950, **18**, 331—343.
  85. **Schwartz, M. M.** The relationship between projective test scoring categories and activity preferences. *Genet. Psychol. Monogr.*, 1952, **46**, 133—181.
  86. **Schwartz, M. M., & Karlin, L.** A new technique for studying the meaning of performance on the Rosenzweig Picture-Frustration Study. *J. consult. Psychol.*, 1954, **18**, 131—134.
  87. **Seward, G. H., & Morrison, L. M.** Personality structure in a common form of colites. *Psychol. Monogr.*, 1951, **65**, 1. Pp. 26.
  88. **Shakow, D., Rodnick, E. H., & Lebeaux, T.** A psychological study of a schizophrenic: Exemplification of a method. *J. abnorm. soc. Psychol.*, 1945, **40**, 154—174.
  89. **Simos, I.** The Picture-Frustration study in the psychiatric situation — preliminary findings. *J. Pers.*, 1950, **18**, 327—330.
  90. **Sinaiko, H. W.** The Rosenzweig Picture-Frustration Study in the selection of department store section managers. *J. appl. Psychol.*, 1949, **33**, 36—42.
  91. **Snedecor, G. W.** *Statistical methods.* Iowa: Iowa State College Press, 1950.
  92. **Sommer, R.** On the Brown adaptation of the Rosenzweig P-F for assessing social attitudes. *J. abnorm. soc. Psychol.*, 1954, **49**, 125—128.
  93. **Symonds, P. M.** Projective techniques. In *Encyclopedia of Psychology*, ed. P. L. Harriman. New York: Philosophical Library, 1946, 583—590.



94. **Symonds, P. M.** Adolescent fantasy. New York: Columbia Univer. Press, 1949.
95. **Symonds, P. M.** Rosenzweig Picture-Frustration Study: A. review. In *Buros* (11), pp. 242—243.
96. **Taylor, M. V., & Taylor, O. M.** Internal consistency of the Group Conformity Rating of the Rosenzweig Picture-Frustration Study. *J. consult. Psychol.*, 1951, **15**, 250—252.
97. **Taylor, M. V.** Internal consistency of the scoring categories of the Rosenzweig Picture-Frustration Study. *J. consult. Psychol.*, 1952, **16**, 149—153.
98. Technical recommendations for psychological tests and diagnostic techniques. *Suppl. Psychol. Bull.*, 1954, **51**, no. 2.
99. **Webster, H.** Rao's multiple discriminant technique applied to three TAT variables. *J. abnorm. soc. Psychol.*, 1952, **47**, 641—648.
100. **Wendland, L. V.** A preliminary study of frustration reactions of the post-poliomyelitic. *J. clin. Psychol.*, 1954. **10**, 236—240.



Norsk revisjon av billed-frustrasjonstesten  
for voksne - utarbeidet og brukt i forbindelse  
med egne undersøkelser

INSTITUTT FOR SAMFUNNSFORSKNING

Nr.....

Arbiensgt. 4, Oslo.

Tlf. 41 35 09.

Dette skjema vil bli behandlet strengt konfidensielt. Dine uttalelser vil bare bli brukt til rent vitenskapelige formål og ditt navn vil aldri bli nevnt.

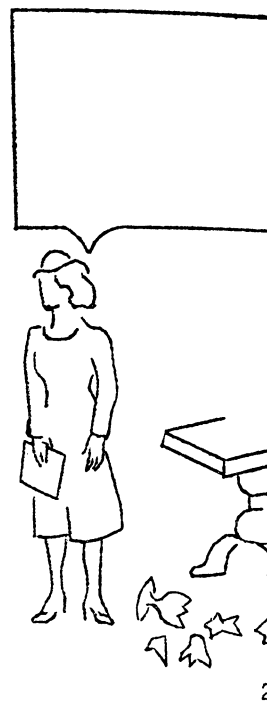
INSTRUKSJON

På hvert av de følgende bilder er det to eller flere mennesker. Det er alltid en person som sier noe til en annen. I den åpne rubrikken skriver du det *aller første som faller deg inn*. Unngå å være morsom. Arbeid så fort du kan.

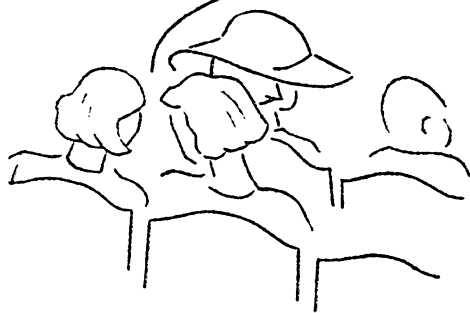
Jeg beklager meget at vi sølte til klærne Deres, men vi gjorde faktisk alt vi kunne for å unngå sølepytten.



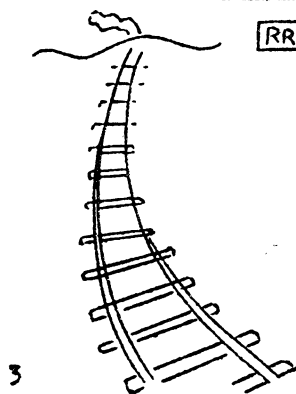
Å, hvor forferdelig. Det var min mors kjæreste vase.



Du ser jo ingen ting.



Det var ergerlig at bilen min fikk motorstopp så du ikke rakk toget.



RR STAT



Nå kommer jeg tilbake med dette fine uret for tredje gang, og det er bare en uke siden jeg kjøpte det.



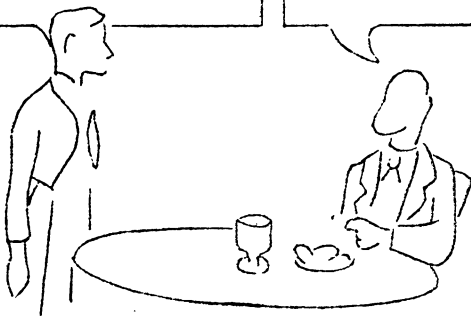
5

Ifølge biblioteksreglementet har De ikke lov til å ta med mer enn 2 bøker om gangen.



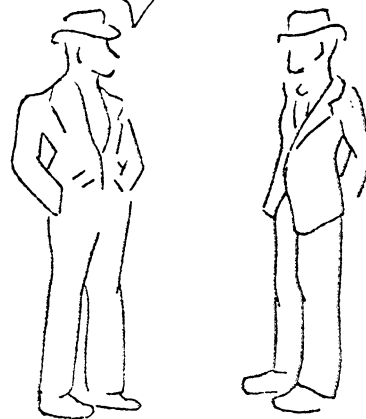
6

Nei, nå syns jeg virkelig De maser vel mye.



7

Jenta di ba meg med på dans i kveld, - hun sa du ikke skulle gå.



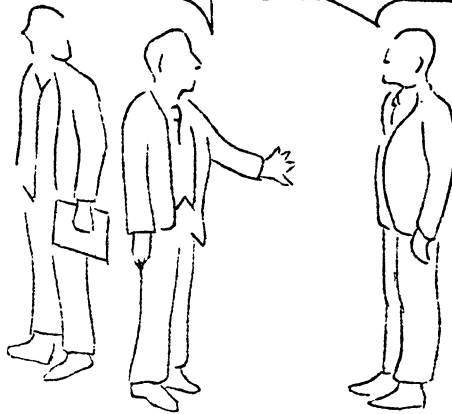
8

Det er vel så at  
De trenger para-  
plyen Deres, men  
De må nok vente  
til sjefen kommer  
i ettermiddag.



9

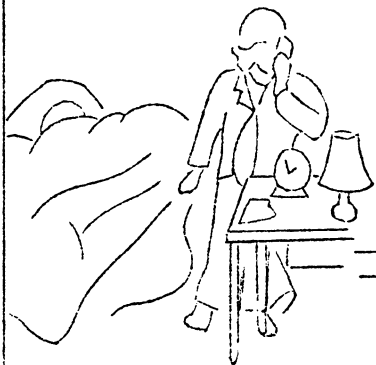
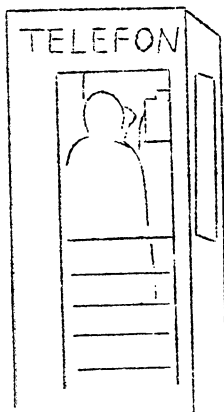
Du lyver, og  
du lyver med  
vilje.



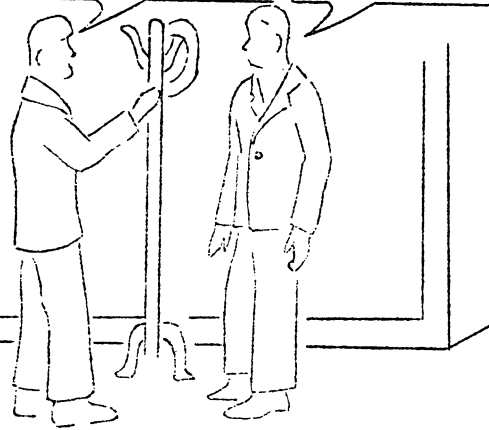
10

(Klokken 2 om  
natten)  
Å, umskyld -  
jeg har nok fått  
feil nummer.

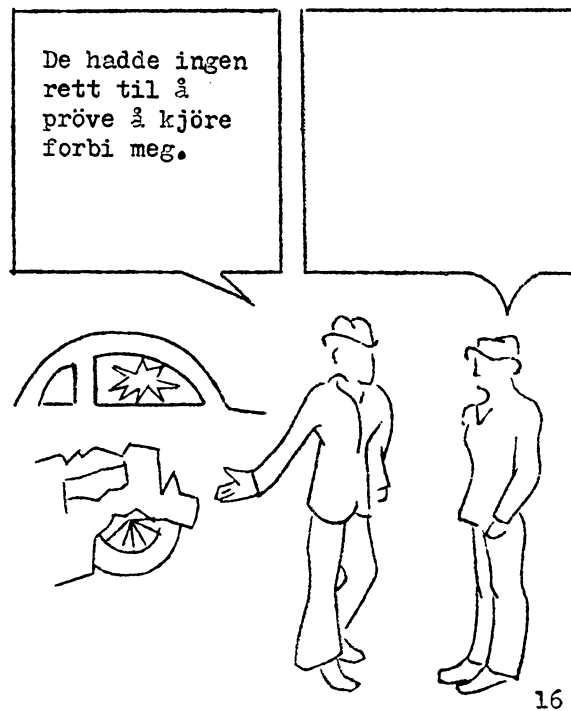
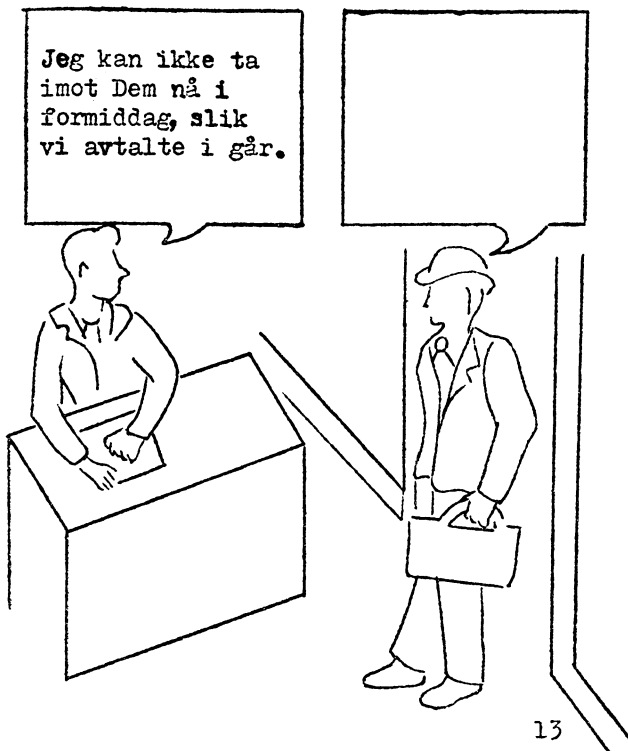
Hvis dette ikke er  
Deres hatt, må Per  
Hansen ha kommet  
til å ta Deres da  
han gikk, og så  
har han latt sin  
egen bli igjen.



11



12

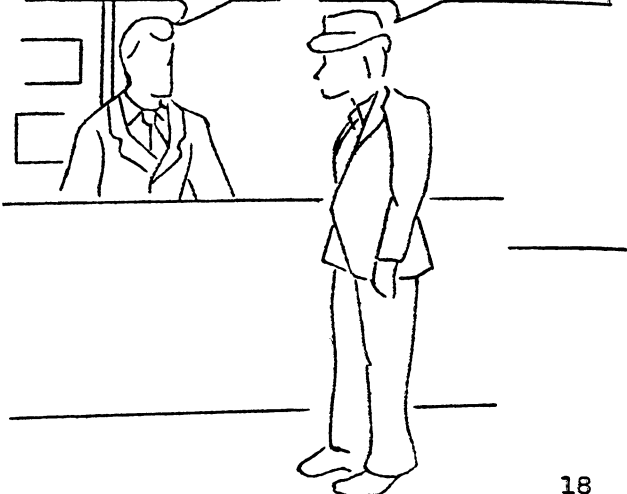


Jo dette er nydelig - at du kommer nettopp nå og sier du har mistet nøklene.



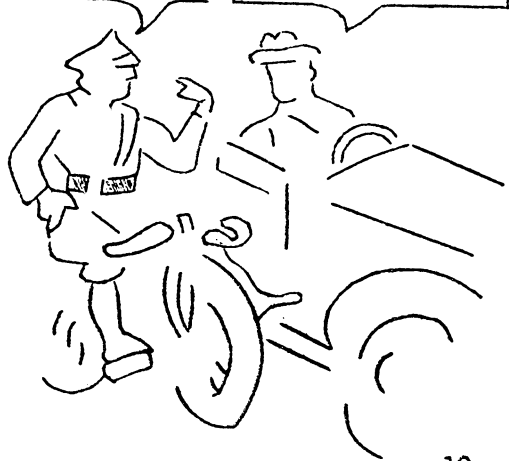
17

Dessverre, nå har vi nettopp solgt den siste.



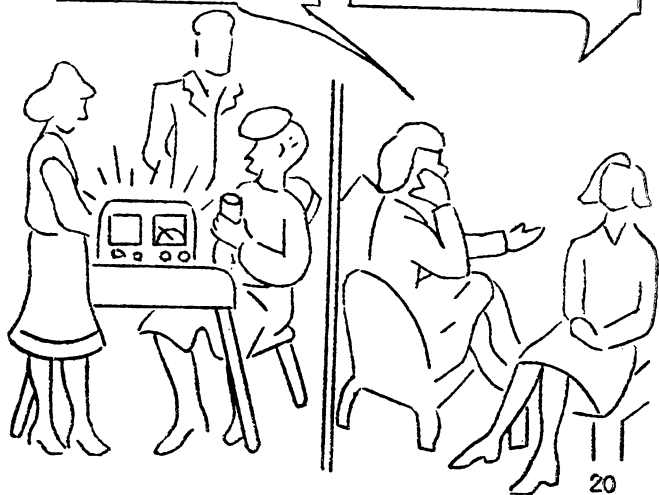
18

Hvor i all verden skal De hen som kjører forbi den skolen med 80 km i timen.



19

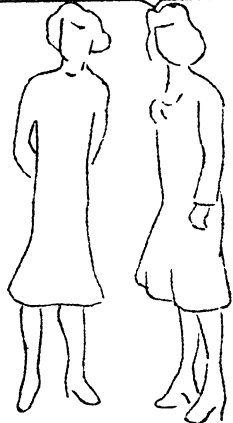
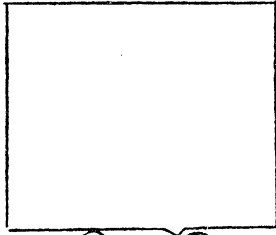
Jeg lurer på hvorfor hun ikke inviterte oss?



20

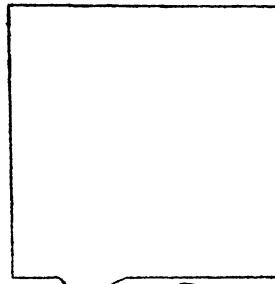
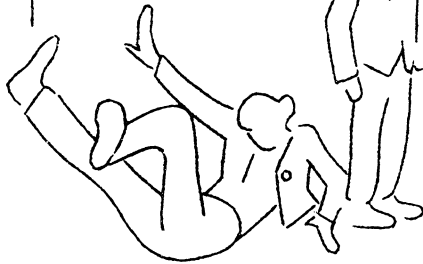


Hun du snakket  
vondt om, var  
utsatt for en  
ulykke i går og  
ligger nå på  
sykehuset.



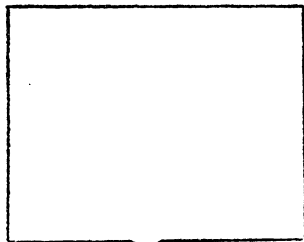
21

Slo De Dem?



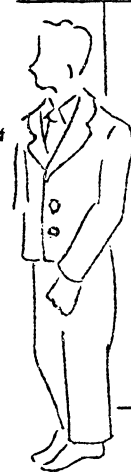
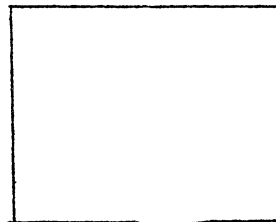
22

Det er Tante.  
Hun ber oss  
vente litt så  
hun kan komme  
ned og si farvel  
enda en gang.



23

Her er den avi-  
sen jeg lånte  
av Dem - ungen  
min rev den dess-  
verre i stykker.



24

*På hvilket grunnlag ga du svarene dine på de foregående sidene?*

- A. Da du skrev replikker, tenkte du deg noengang selv i samme situasjon og hva du ville ha sagt i situasjonen? (Sett sirkel om det tallet som står under det svar du synes passer.)

| Alltid | Ofte | Iblant | Sjelden | Aldri |
|--------|------|--------|---------|-------|
| 1      | 2    | 3      | 4       | 5     |

- B. Da du skrev replikker, svarte du noengang det du antok du selv faktisk ville ha sagt i samme situasjon?

| Aldri | Sjelden | Iblant | Ofte | Alltid |
|-------|---------|--------|------|--------|
| 1     | 2       | 3      | 4    | 5      |

- C. Da du skrev replikker, svarte du noengang det du mente du *burde* ha sagt i samme situasjon, uten tanke på hvorledes du faktisk mest sannsynlig ville ha oppført deg i situasjonen?

| Aldri | Sjelden | Iblant | Ofte | Alltid |
|-------|---------|--------|------|--------|
| 1     | 2       | 3      | 4    | 5      |

Navn: ..... Født:.....

## *Appendiks II*

Norsk revisjon av skåreskjema  
til billed-frustrasjonstesten — utarbeidet  
i forbindelse med egne undersøkelser

ROSENZWEIGS B-F-TEST  
Skåreskjema

Nr. ....

Navn: ..... Alder:.....

Dato: ..... Adm. av:..... Adm. tid:.....

| Leddskårer |     |     | Skåreprofil |        |     |     |   |
|------------|-----|-----|-------------|--------|-----|-----|---|
| H-D        | E-D | B-P |             | H-D    | E-D | B-P | Σ |
| 1          |     |     | Ex          | 1. del |     |     |   |
| 2          |     |     |             | 2. del |     |     |   |
| 3          |     |     | In          | 1. del |     |     |   |
| 4          |     |     |             | 2. del |     |     |   |
| 5          |     |     | Im          | 1. del |     |     |   |
| 6          |     |     |             | 2. del |     |     |   |
| 7          |     |     | Σ           | 1. del |     |     |   |
| 8          |     |     |             | 2. del |     |     |   |
| 9          |     |     |             |        |     |     |   |
| 10         |     |     |             |        |     |     |   |
| 11         |     |     |             |        |     |     |   |
| 12         |     |     |             |        |     |     |   |

S-E skårer og mønstre

|    |              |        |
|----|--------------|--------|
| 13 | <u>E</u>     | E — E  |
| 14 |              |        |
| 15 | <u>I</u>     | I — I  |
| 16 |              |        |
| 17 | <u>E + I</u> | Im + I |
| 18 |              |        |

Kommentarer: