

# Kartleggingsverktøy

*En pedagogisk analyse av noen aktuelle verktøy som brukes i Norge i forhold til kartlegging av norskspråklige ferdigheter hos språklige minoriteter.*

**Karoline Sem Nilsen**



Masteroppgave ved Pedagogisk forskningsinstitutt

UNIVERSITETET I OSLO

Juni 2008

---

## Innhold

|   |           |
|---|-----------|
| <b>INNHold</b> .....  | <b>2</b>  |
| <b>FORORD</b> .....   | <b>4</b>  |
| SAMMENDRAG AV MASTEROPPGAVEN .....  | 5         |
| <b>1. INNLEDNING</b> .....  | <b>10</b> |
| <b>2. BAKGRUNN FOR ANALYSEN</b> .....   | <b>12</b> |
| <b>3. METODISKE REFLEKSJONER</b> .....  | <b>16</b> |
| 3.1 METODEGRUNNLAG .....  | 17        |
| <b>4. TEORETISK DEL</b> .....   | <b>20</b> |
| <b>5. EMPIRISK DEL</b> .....  | <b>24</b> |
| 5.1 PSYKOLOGISK TESTING.....  | 25        |
| 5.2 WISC III - ET VERKTØY FOR KARTLEGGING AV INTELLEKTUELL KAPASITET HOS BARN. .... | 31        |
| 5.2.1 Hva måler testen .....  | 33        |
| 5.2.2 Testens kulturelle kontekst og opprinnelse.....                               | 34        |
| 5.2.3 Testens normeringsutvalg .....  | 34        |
| <b>6. SPRÅK 6 -16</b> .....   | <b>37</b> |
| 6.1.1 Hva måler testen .....  | 37        |
| 6.1.2 Testens kulturelle kontekst og opprinnelse.....                               | 39        |
| 6.1.3 Testens normeringsutvalg .....  | 40        |
| 6.1.4 Ordvalg i testen.....   | 41        |
| <b>7. TEST FOR RECEPTION OF GRAMMAR 2ED – TROG 2</b> .....                          | <b>43</b> |
| 7.1.1 Hva måler testen .....  | 43        |
| 7.1.2 Testens kulturelle kontekst og opprinnelse.....                               | 44        |

---

|            |   |           |
|------------|---|-----------|
| 7.1.3      | <i>Testens normeringsutvalg</i> .....                       | 44        |
| 7.1.4      | <i>Ordvalg i testen</i> .....                               | 44        |
| <b>8.</b>  | <b>THE BRITISH PICTURE VOCABULARY SCALE - BVPS II</b> ..... | <b>46</b> |
| 8.1.1      | <i>Hva måler testen?</i> .....                              | 46        |
| 8.1.2      | <i>Testens kulturelle kontekst og opprinnelse</i> .....     | 47        |
| 8.1.3      | <i>Testens normeringsutvalg</i> .....                       | 47        |
| 8.1.4      | <i>Ordvalg i testen</i> .....                               | 47        |
| <b>9.</b>  | <b>SAMMENLIGNING</b> .....                                  | <b>49</b> |
| 9.1.2      | <i>Bildene i TROG 2 og BPVS II</i> .....                    | 51        |
| <b>10.</b> | <b>REFLEKSJONER OVER FUNN FRA ANALYSEN</b> .....            | <b>52</b> |
| <b>11.</b> | <b>PEDAGOGISKE IMPLIKASJONER</b> .....                      | <b>55</b> |
| 11.1.1     | <i>Hva er alternativet?</i> .....                           | 58        |
| <b>12.</b> | <b>AVSLUTTENDE KOMMENTARER</b> .....                        | <b>60</b> |
|            | <b>KILDELISTE</b> .....                                     | <b>62</b> |

## Forord

*“Tests not accompanied by detailed data on their construction, validation, uses, and limitations should be suspect.”*

- Oscar K. Buros (Buros i Sattler 2001:1).

Denne oppgaven tar utgangspunkt i tre ulike tester, som brukes til kartlegging av norsk språklige ferdigheter. Minoritetsspråklige elevers norsk språklige ferdigheter er svært viktig for deres skolegang. Adekvat kartlegging kan ha betydning for hva slags hjelpetiltak som settes i gang rundt eleven, dersom eleven skulle ha problemer i forhold til språklig utvikling. Flerkulturell pedagogikk har vært og er fremdeles et tema som jeg har interessert meg siden jeg begynte på universitetet. Derfor falt det naturlig å velge å skrive masteroppgave innenfor dette området. Det har vært en lang og lærerik prosess. Særlig har arbeidet med masteroppgaven vært spennende fordi jeg har fått mulighet til å fordype meg i et spesifikt område.

Arbeidet med min materoppgave i pedagogikk, ved Institutt for Pedagogisk Forskning, Universitetet i Oslo, er ved veis ende. Det er på mange måter godt å bli ferdig, men det er også rart å vite at jeg snart skal ta skrittet ut i arbeidslivet, hvor kanskje ikke alt stemmer like godt overens med det som står i læreboka. Jeg vil gjerne takke Kamil Øzerk for lærerike og interessante samtaler og veiledning, samt oppmuntrende ord. Videre vil jeg takke familie og venner for god støtte og forståelse. Jeg vil også takke studievenninner for hyggelige kaffepauser, både med og uten faglig innhold.

Karoline Sem Nilsen

Juni 2008

---

## SAMMENDRAG AV MASTEROPPGAVEN

**TITTEL:**

Kartleggingsverktøy.

**UNDERTITTEL:**

En pedagogisk analyse av noen aktuelle verktøy som brukes i Norge i forhold til kartlegging av norskspråklige ferdigheter hos språklige minoriteter.

**AV:**

Karoline Sem Nilsen.

**EKSAMEN:**

Masterstudiet i pedagogikk.

Studieretning: Pedagogisk psykologisk rådgivning.

**SEMESTER:**

Våren 2008.

**STIKKORD:**

Pedagogisk psykologi.

Flerkulturell pedagogikk.

Minoritetsspråklige elever.

Testing.

**Problemområde:**

Denne oppgaven har kartlegging av norskspråklige ferdigheter hos minoritetsspråklige elever som hovedområde. Oppgaven tar utgangspunkt i tre ulike kartleggingsverktøy som undersøker språkutvikling hos barnet. Disse testene har blitt analysert. Oppgavens problemstilling er følgende: I hvilken grad er TROG 2, BPVS II og Språk 6 – 16 gode kartleggingsverktøy for mål av norskspråklig utvikling hos minoritetsspråklige elever i Norge?

Minoritetsspråklige elever er en stadig økende gruppe i skolen, dette kan i mange tilfeller føre til nye problemstillinger i skolen. Norsk språklig utvikling hos denne gruppen er viktig i forhold til skolesuksess. Derfor er kartlegging av språk hos denne elevgruppen, være svært viktig. Dette krever at det finnes adekvate kartleggingsverktøy, slik at eleven kartlegges riktig, og dermed får den hjelp som trengs når det gjelder barnets språkutvikling. De testene som har blitt analysert i denne oppgaven, er verken tiltenkt kartlegging av språklig ferdigheter hos minoritetsspråklige elever, samtidig som det mangler normeringsgrunnlag til denne elevgruppen. Dette kan ses på som en svakhet ved samtlige av testene.

Det er fem punkter som har vært av særlig interesse i analysedelen: Hva måler testen?, Testens kulturelle kontekst og opprinnelse, Testens normeringsutvalg, Ordvalg i testen og Bildene i testen. Dette ses i sammenheng med testens reliabilitet, validitet, operasjonalisering, normering/standardisering og testens representativitet. Dette er begreper som går inn i hverandre, og som har betydning for barnets resultater på testen, men også når det gjelder påliteligheten og gyldigheten av testresultatene.

**Metode**

Denne oppgaven er basert på ulik metodebruk. Metodene er som følger; en litteraturstudie samt en empirisk studie av utvalgte kartleggingsverktøy. Litteraturen er hovedsaklig et utvalg av relevant pensum, samt manualene til testene og selve testen. Den norske manualen til TROG 2 er kun en foreløpig utgave, da den i skrivende stund fremdeles er under utvikling. Den engelske manualen er brukt som utgangspunkt for diskusjonen rundt BPVS II, samt det norske skåringsarket.

**Resultater**

Det jeg har kommet frem til gjennom analysen er at de utvalgte testene ikke er det mest hensiktsmessige kartleggingsverktøy for kartlegging av språklig ferdigheter hos minoritetsspråklige elever i Norge. Det viser seg også at det ikke finnes mange alternativer for kartlegging av norskspråklige ferdigheter hos minoritetsspråklige elever. Jeg har også kommet frem at det kan være nødvendig å utvikle parallelle kartleggingsverktøy på barnets morsmål, slik kan man undersøke om barnet har språkvansker i begge språk, eller om det er relaterte norskspråklige vansker/mangler.







# 1. Innledning

Andelen minoritetsspråklige elever blir stadig større i norsk skole. Dette vil i enkelte tilfeller føre til nye problemstillinger i skolen. Bare i Oslo skolen viser det seg at 48 av grunnskolene har over 50 % elever med et annet morsmål enn norsk (Aften 08.01 2008). Andelen innvandrerbarn utgjør i dag omlag 9 % av den totale folkemengden (SSB 2007). Mange av disse har minoritetsspråklig bakgrunn. Minoritetsspråklig bakgrunn kan defineres som individer med et annet språk, og annen kulturbakgrunn enn norsk, svensk, dansk og engelsk (Kunnskapsdepartementet 2007a). Det vil si at disse barna vokser opp med et annet morsmål enn norsk. Det å kunne flere språk anses som positivt, men disse barna har som oftest en større utfordring i forhold til norskspråklig utvikling. Barna skal tilegne seg basisferdigheter som lese, skrive og regne, samtidig som de møter et nytt språk. De møter altså en dobbel utfordring når de begynner på skolen, som for mange kan bety at de henger etter rent faglig (Bakken 2007). En evaluering av spesialundervisningen i Oslo skolen (Nordahl & Overland 1998) viser at andelen tospråklige elever er overrepresentert i forhold til spesialpedagogisk hjelp. Årsakene til dette er sammensatt bilde, men at disse elevene har større utfordringer vedrørende språk er et faktum. Både i forhold til norskspråklig utvikling, og morsmålsutvikling. Det er en viss overensstemmelse i fagfeltet at tospråklige elever i mange tilfeller vil kunne dra nytte av å beherske to språk flytende (Cummins 1979; Øzerk 2003). Dette betinger at språkene utvikler seg på en tilfredsstillende måte. En ønskelig språkutvikling kan betegnes som additiv tospråklig utvikling. Der barnets andrespråksutvikling ikke skjer på beskosting av morsmålet (Cummins 1979, Cummins 2000; Øzerk 2003). En slik utvikling kan som sagt være vanskelig for et flertall minoritetsspråklige barn, da de ofte møter norsk språk på et senere tidspunkt. Mange strever derfor med tilegnelse av norsk språk. Dette kan være en mulig årsak til at denne elevgruppen er overrepresentert i forhold til spesialpedagogiske tiltak.

Dersom minoritetsspråklige barn strever med norskspråklig utvikling, kan det være nødvendig med en adekvat kartlegging, for å avdekke vanskene til barnet. Reflekterer det kartleggingsmaterialet som foreligger per dags dato, et mål av minoritetsspråklige elevers språklig utvikling? Utgangspunktet for denne masteroppgaven er en analyse av utvalgte kartleggingsmateriale som måler språklig utvikling. Jeg ønsker å se på kartleggingsmateriale som foreligger i forhold til kartlegging av språklige ferdigheter, hos minoritetsspråklige

elever i Norge. Følgende tester vil bli diskutert gjennomgående: TROG 2, Test for Reception of Grammar (Bishop 1989), BPVS II, The British Picture Vocabulary Scale 2.ed (Dunn & Dunn mfl 1997) og Språk 6-16 (Ottem & Frost 2005). Problemstillingen lyder som følger: ”I hvilken grad er TROG 2, BPVS II og Språk 6 – 16 gode kartleggingsverktøy for mål av norsk språklig utvikling hos minoritetsspråklige elever i Norge? For å kunne besvare dette, har jeg gjennomført en deskriptiv analyse av testene.

Oppbyggingen av oppgaven er som følgende, oppgaven starter med bakgrunnen for valg av problemstilling. Deretter tas det opp en rekke metodiske refleksjoner rundt temaet. Videre følger en teoretisk del som tar utgangspunkt i generell testutvikling. Sentrale begreper i dette kapitlet vil være reliabilitet, validitet, operasjonalisering, normering og representativitet. Neste kapitel tar i hovedsak opp psykologisk testing, samt diskusjonen rundt bruk av tester og kontroverser rundt dette. WISC III, en ofte brukt intelligens test blir gjennomgått, for å vise et eksempel på en test som har vært gjennom mye kritikk for bruk på minoritetsspråklige elever. Deretter gjennomgås testene som er under analyse. Så følger refleksjonene som er gjort ovenfor funnene fra analysen og pedagogiske implikasjoner som kan komme som følge av dette.

## 2. Bakgrunn for analysen

Bakgrunnen for analysen tar utgangspunkt i en evaluering utført av Rambøll Management og Kunnskapsdepartementets strategiplan: "Likeverdig opplæring i praksis, 2007- 2009".

Utdanningsdirektoratet gjennomførte en evaluering av praktiseringen av norsk som andre språk for språklige minoriteter i grunnskolen. Evalueringen ble gjennomført av Rambøll Management. Formålet var blant annet å fremskaffe kunnskap om hvordan kommunene praktiserte særskilt norsk undervisning for språklige minoriteter (Rambøll Management 2006). Det fremkommer av rapporten at enkelte av kommunene mangler retningslinjer i forhold til hvilke barn som skal ha særskilt norsk opplæring. Av særlig interesse er rapportens kommentarer vedrørende skolenes kartleggingsmetoder for å avdekke hvilke elever som trenger særskilt norsk opplæring. Denne beskrives som utilstrekkelig og mangelfull. Kartleggingsmetodene varierer fra kommune til kommune. Samtidig foreligger det en rekke ulike kartleggingsverktøy som skolene selv har utviklet for å bidra i evalueringen av språkutviklingen til minoritetsspråklige elever (Rambøll Management 2006). Dette kan ses på som bekymringsfullt dersom det forekommer en differensiering av minoritetsspråklige elever i skolene, når det gjelder hvem som får særskilt norsk opplæring. Det konstateres i rapporten fra Rambøll Management at det er et behov for standardiserte kartleggingsverktøy som avdekker minoritetsspråklige elevers norsk språklige ferdigheter (2006).

Kunnskapsdepartementets strategiplan: "Likeverdig opplæring i praksis!" har som formål å øke læringsutbytte og deltakelsen av språklige minoriteter i barnhage, skole og utdanning 2007- 2009. Strategiplanen inneholder en rekke punkter som skal jevne ut forskjellene mellom minoritetsspråklige barn i barnehage, grunnskole og videregående skole. Videre tar strategiplanen sikte på å bidra til økt deltagelse blant språklige minoriteter i forhold til høyere utdanning, samt deltakelse i arbeidsliv og samfunnsliv (Kunnskapsdepartementet 2007).

Tiltak 13 i strategiplanen lyder som følger: "*Kartlegging og utredning av elevens ferdigheter og behov*".(30: 2007). Utdanningsdirektoratet vil i dette anseende sørge for at det foreligger relevante utredningsverktøy for språklige minoriteter. Som prosjektleder har utdanningsdirektoratet igangsatt Bredtvet og Torshov kompetansesenter, i forhold til dette:

---

*”- utarbeide og prøve ut metoder for observasjons- og kartleggingsmateriell i forbindelse med utredning av elever med behov for spesialpedagogisk hjelp.” (Kunnskapsdepartementet 30: 2007).*

Prosjektet til Bredtvet tar utgangspunkt i tre tester:

- 1) TROG 2, Test for Reception of Grammar (Bishop 1989)
- 2) BPVS II, The British Picture Vocabulary Scale 2.ed (Dunn & Dunn mfl 1997) og
- 3) Språk 6-16 (Ottem & Frost 2005).

Analysen i denne oppgaven tar også utgangspunkt i disse testene, men må ses som uavhengig av prosjektet til Utdanningsdirektoratet. Prosjektet til Utdanningsdirektoratet har til hensikt å videreutvikle disse testene. Det ser ut til at tiltak 13 er et resultat av evalueringen fra Rambøll Management, hvor det påpekes manglende retningslinjer og standardiserte kartleggingsverktøy for minoritetsspråklige elever.

Analysen tar fortrinnsvis utgangspunkt i de overnevnte testene. Fokusområdene vil være testenes sterke og positive sider med tanke på minoritetsspråklige elever. Det vil stilles en rekke spørsmål vedrørende hver av testene, som forsøkes besvares gjennom en analyse. Et vesentlig moment vil være hva testen tar sikte på å måle, og i hvilken grad testen lever opp til dette. Testenes kulturelle kontekst vil være av betydning, med tanke på at minoritetsspråklige barn ofte har en annen kulturell bakgrunn enn majoritetsspråklige barn. Det er særlig fem begreper som er viktig i analysedelen, begrepene er som følger: reliabilitet, validitet, operasjonalisering, normering/standardisering og representativitet. Disse ses sammen med hva testen måler, testens kulturelle kontekst og opprinnelse, testens normeringsutvalg samt ordvalg og bilder i testen. Hvorvidt minoritetsspråklige elever er tatt med under utviklingen av en test, er særdeles viktig med tanke på videre bruk på denne elevgruppen.

Opplæringsloven, § 1-2 femte ledd lyder som følgende:

*”Opplæringa skal tilpassast evnene og føresetnadene hjå den enkelte eleven, lærlingen og lære kandidaten”.*

Dersom eleven ikke får tilfredsstillende utbytte av den ordinære opplæringen, har eleven i henhold til Opplæringsloven § 5-1 rett til spesialundervisning. Skal eleven ha spesialpedagogisk hjelp forutsettes det at det foreligger en sakkyndig vurdering. En sakkyndig vurdering skal i følge opplæringsloven § 5-2 bestå av en vurdering av elevens særlige behov (Opplæringsloven, 1998). Dette innebærer at det blir tatt standpunkt til følgende: elevens lærevansker, realistiske opplæringsmål for eleven, hvorvidt eleven kan hjelpes innenfor tilbudet om tilpasset opplæring og i hvilken grad opplæringstilbudet som gis er forsvarlig (Opplæringsloven § 5-3, 1998). En sakkyndig vurdering vil i mange tilfeller innebære en form for testing av eleven. Dette har på mange måter ført til etterspørsel av ulike tester i forhold til barn i skolen. Dessuten blir ikke bare elevene målt i forhold til individuelle standardiserte og normerte tester, men også i forhold til nasjonale mål. Det finnes nasjonale kartleggingsprøver i både lesing, norsk, engelsk og matte som elevene testes i. Videre er det også internasjonale sammenligningstester som TIMMS OG PISA. Dette begynner å bli en viktig del av utdanningsdebatten og elevenes læring på skolen. Det viser seg at det særlig er behov for nyutviklende testinstrumenter, for å få en adekvat kartlegging av minoritetsspråklige elevers ferdigheter. Utvikling av nye tester er tidkrevende og spesielt standardiserte tester. Standardiserte tester krever en normeringsgruppe, det vil si en sammenligningsgruppe, som individets resultater ses opp i mot.

Dersom testen ikke er utviklet i Norge, trenger den en omfattende videreutvikling for å kunne brukes i Norge. Det skal blant annet foretas en normering av testen, for å kunne gi et optimalt sammenligningsgrunnlag i forhold til norske elever. Dette vil sikre at testen gir reliable og valide resultater. Testen gjennomgår også ofte en oversettelsesprosess. Det viser seg at normeringen av en test ofte tar lengre tid enn selve oversettelsesprosessen. Fokus på ordvalg og kulturelle bidrag i testene får omfattende oppmerksomhet, og man kan spørre om det i like stor grad legges vekt på testens psykometriske egenskaper, og i hvilken grad disse ivaretas i en oversettelsesprosess. Denne oppgaven ønsker å sette fokus på disse nyutviklende testene, spesielt i forhold til når disse tas i bruk på språklige minoriteter, og vise hva slags fordeler og ulemper som kommer til syne, under overflaten til de utvalgte testene.

Skolene har et ansvar for elever med skolefaglig, personlig eller sosiale vansker, som krever at det foreligger kartleggingsprosedyrer. Utviklingen av utredningsverktøy bør derfor ta hensyn til ulike elevgrupper i skolen, slik at kartleggingen foregår optimalt i forhold til enkeltelevers vansker. Slik unngår man unødig synsing i forhold til elevens vansker på

---

skolen. Samtidig kan egnede tester hjelpe sakkyndige og lærere å sette ord på elevens vansker og utvikle tilpassende opplæringstilbud deretter.

### 3. Metodiske refleksjoner

*”En av de viktigste begrunnelsene for bruk av metode er å sikre holdbarheten av de påstander vi kommer med.” (Kvernbekk i Lund 2002).*

Denne oppgaven er basert på ulike metodebruk. Metodene er som følger; en litteraturstudie samt en empirisk studie av utvalgte kartleggingsverktøy. Årsaken til disse valgene er som nevnt i sitatet, å sikre holdbarheten i de påstandene som kommer frem i oppgaven. Metodene er også valgt fordi de utfyller hverandre, slik at oppgavens problemstilling lar seg besvare. En slik kombinasjon av metodebruk, kalles gjerne triangulering. Dette innebærer at man bruker ulike metoder for sikre et bestemt fenomen (Alvesson & Sköldberg 1994). Hva som er valid forskning vil endres over tid, særlig i forhold til hvilket vitenskapsteoretisk syn en har som utgangspunkt. Popper har uttrykt at vitenskapens mål er å oppnå sannhet. For å vise dette er det viktig å kunne skille vitenskapelig aktiviteter fra ikke- vitenskapelig aktiviteter (Kvernbekk i Lund 2002). Popper forslår falsifisering som et slikt skille, dette kalles demarkasjonskriteriet, og går ut på å avkrefte en påstand, altså at finne ut hvorvidt påstanden er uforenelig med observasjoner/ testresultater i virkeligheten. Forskning vil da dreie seg om å avkrefte tidligere teorier. Rent teoretisk kan det altså fremskaffes empiri på ulike sider av et forskningsfelt som avkrefter teoriene i ulike forskningstradisjoner. Derfor kan Poppers måte og drive vitenskapelig aktivitet, være noe motstridende, fordi det nærmest er umulig å oppnå sannhet (Hjardemaal i Kleven 2002). Det kan derfor være vanskelig å komme frem til sannhet, når det stadig kommer frem nye forskningsresultater som motbeviser tidligere forskning. Det vil likevel være et forsøk å komme nærmere sannheten. Dersom en påstand stemmer med virkeligheten er den i følge korrespondanseteorien sann (ibid). Årsaken til at dette nevnes er for å påpeke at det alltid vil være mulig å komme frem til et annet synspunkt som er avvikende fra det som i skrivende stund anses som riktig.

I fagene pedagogikk og psykologi er det vesentlig å nevne at det finnes et såkalt målingsproblem i forhold til innhenting av data. Dette betyr at man opererer med begreper som vanskelig lar seg måle. De fenomenene som ønskes studert, er ofte ikke direkte observerbare. Dette betyr at empirisk forskning i disse fagene kan være vanskelig. Empirisk forskning drives til tross for denne problematikken, fordi man velger å tillegge de ikke observerbare fenomenene visse indikatorer som peker i retning av det fenomenet som ønskes studert (Kleven 2002). Jo bedre dette gjøres, desto mer relevant og presis vil



---

datainnsamlingen være. I de testene som denne oppgaven har som utgangspunkt, har testutviklerne operasjonalisert begrepene som skal måles så godt som det lar seg gjøre. Utviklingen av testene er langt mer komplisert enn beskrevet her, og det er foretatt ulike operasjonsluseringer innenfor hver enkelt deltest. Hvordan disse begrepene er operasjonalisert, har betydning for testens signifikans. Spesielt i forbindelse med hvilken barnegruppe testen er utviklet for (Kleven 2002).

Psykometri kvalitetssikrer psykologiske tester. Psykometri dreier seg om å måle ulike psykologiske tester, og ikke hva selve testen måler. For å gjennomføre dette bruker man en rekke statistiske metoder for å analysere en utvalgt test. Dette gjøres blant annet gjennom korrelasjonsanalyser, og andre statistiske metoder. Innenfor psykometri er det utviklet metoder for å etterprøve de ulike indikatorene, og hvordan disse synes og representere samme begrep (Kleven 2002). Dette vil gi et helhetsbilde av testen, og dermed vise til hvorvidt dette er en god test. I denne sammenheng blir validitet og reliabilitet to sentrale begreper.

### 3.1 Metodegrunnlag

Metodegrunnlaget i denne oppgaven baseres blant annet på en analyse av TROG, Test for Reception of Grammar 2.ed (Bishop 1989), BPVS II, The British Picture Vocabulary Scale 2.ed (Dunn & Dunn mfl 1997) og Språk 6-16 (Ottem & Frost 2005). Fokuseringsområdene i analysen er delt i to deler, en teoretisk og en empirisk del. I den teoretiske delen vil jeg først og fremst beskrive generelle betingelser for utviklingen av psykometriske instrumenter.

Kildegrunnlaget baseres hovedsakelig på utvalgt litteratur som beskriver generell testutvikling, psykometri og statiske verktøy som brukes under utviklingen av ulike tester.

Det er særlig fire momenter som er vesentlig i denne delen:

- a) Testens validitet
- b) Testens reliabilitet
- c) Testens representativitet og
- d) Testens normering.

Diskusjonen vil fortrinnsvis ta utgangspunkt i disse begrepene, og vil derfor diskuteres ut fra generelle vilkår. Rent spesifikt i forhold hvordan dette er tatt hensyn til i de utvalgte testene vil tas opp under de respektive kapitlene til hver enkelt test som er analysert. Et gjennomgående trekk i disse kapitlene vil være i hvilken grad minoritetsspråklige elever er tatt hensyn til under utviklingen av testen. Dette vil ha stor betydning for testens reliabilitet, validitet, representativitet og normering i forhold til bruk av testen til kartlegging av minoritetsspråklige elevers norskspråklige utvikling. Dette gjelder særlig i sammenheng med norsk oversetting, normering og standardisering, men også testens opprinnelse. Har testens originalutvikler hatt språklige minoriteter som målgruppe, da testen ble utviklet?

Den empiriske delen består av en analyse av innholdet i de tre testinstrumentene. Analysen tar for seg den kulturelle konteksten og opprinnelsen til testen. Det vil si at fokusområdet vil være testens kulturelle innhold, elevenes erfaringsverden sett opp i mot testens kulturelle referanserammer. Videre vil analysen ta for seg hva testen tar sikte på å måle, dette ses i sammenheng med testenes reliabilitet og validitet, samt testens egnethet. Testen kan ha som utgangspunkt å måle grammatiske ferdigheter, mens det samtidig er andre grunnleggende ferdigheter som for eksempel et godt og velutviklet vokabular for å oppnå full skåre på oppgaven. Hvorvidt testene er normert er særdeles viktig i analysen, fordi dette viser til testenes reliabilitet og validitet, og i hvilken grad man kan stole på testresultatene. Dersom testen er normert, er det vesentlig å stille spørsmål om hvor normeringen er foretatt, og hvilke andel fra populasjonen som er tatt med. Kildene i den empiriske delen, vil først og fremst være manualen, og selve testprotokollen. Det betyr at disse kildene vil forklare og utfylle problemstillingen i oppgaven.

### **Reliabilitet og validitet i forhold til analysen.**

Hvordan sikre at det trekkes valide og reliable slutninger i analysen? Dette spørsmålet er aktuelt for forskere. Det er verdifullt å stille samme spørsmål ved denne oppgaven, i den grad man kan komme frem til valide og reliable slutninger, basert på teoretisk litteratur bidrag. Fallgruven i oppgaven, kan være at man lett kan foreta ubegrunnede slutninger. Begrunnelsene for slutningene er derfor basert på et nøye utvalgt teoretisk bidrag. Slik blir det mer plausibelt å sikre et valid og reliabelt resultat. Oppgaven vil i det minste kunne stille noen spørsmål angående testutvikling, samt vise til en del punkter som er aktuelt å se ytterligere på under utvikling av nye tester eller videreutvikling av tester fra andre

land/regioner. Å være kildekritisk er derfor en forutsetning for å sikre påstandene man kommer med.

Både reliabilitet og validitet har betydning for forskning. Analysen i denne oppgaven, bør derfor både være gyldig og pålitelig, det vil si valid og reliabel.

## 4. Teoretisk del

Denne delen vil hovedsakelig dreie seg som generell testutvikling. Sentrale begreper vil være reliabilitet, validitet, operasjonalisering, normering/standardisering og representativitet. De statistiske konseptene som er underliggende i test, er vesentlig å ha et innblikk i for å kunne se hva som ligger bak de resultatene som fremkommer fra testutredningen, og ikke minst for å kunne foreta en riktig tolkning av individets resultater.

Det brukes en rekke statistiske metoder for å analysere en utvalgt test. Klassisk testteori har som utgangspunkt at hvert individ har en "ekte" testskåre innenfor en test. Oppnådd testskåre representerer da den "ekte" skåren samt ulik mengde feil i målingen (Walsh & Betz 1995). Jo mindre denne feil, desto bedre reliabilitet har testen.

Å utvikle en test er en avansert prosess. For å sikre best mulig kvalitet i sluttproduktet, er det hensiktsmessig å følge visse metoder innenfor testkonstruksjon. Som nevnt er reliabilitet og validitet viktige begreper, men disse begrepene fremkommer ikke før testen er ferdigutviklet (Walsh & Betz 1995). En testkonstruksjon bør inneholde følgende fem trinn:

1. Riktig definisjon av begrepet som skal måles - Operasjonalisering av begrepet.
2. Utvikling av items/oppgaver som er logisk relatert til begrepet som ønskes målt.
3. Administrere items/oppgavene på et stort utvalg individer.
4. Justering av de opprinnelige items/oppgavene. Gjennom administreringen i punkt tre, vil det kunne fremkomme ulike items som lar seg eliminere bort, eller andre items som bør tilføres.
5. Administrering av testen til en ny gruppe individer (Walsh & Betz 1995).

Item analyse er en viktig del under konstruksjonen av en test. Denne analysen avgjør blant annet hvilken vanskelighetsgrad de ulike itemene/oppgavene skal ha, og fastsetter hvilke items som skal være i testen. De utvalgte items kan i noen tilfeller vise seg å ikke måle det som er tenkt at det skal måle, dermed må disse diskrimineres (ibid).

### **Reliabilitet**

Reliabilitet kan defineres som pålitelighet, støhet, stabilitet (Bø & Helle 2002).

---

*”I samfunnsvitenskapelige undersøkelser viser reliabilitet til målingspåliteligheten av en egenskap uten å si noe om hva som måles”. (Bøe & Helle 2002: 212). ’*

Det dreier seg altså om hvorvidt gjentatte målinger med samme instrument gir samme resultat (Ringdal 2007). Psykometriske instrumenter vil aldri kunne gi resultater som er fullstendig reliable. Det vil alltid forekomme en grad av usikkerhet rundt en testskåre. Det er i prinsippet ikke mulig å teste et individ to ganger uten at den første testingen påvirker testing nummer to. Et pedagogisk eller psykologisk mål må kunne sies å være reliabel hvis målingen har noen form for betydning. Reliabilitet refererer til konsistens på ulike plan. Først og fremst bør målingen være i samsvar med seg selv, dette kalles indre reliabilitet. Videre må reliabiliteten være konsistens over tid (test- retest reliabilitet). Det vil si at det som måles er stabilt over tid, og dersom det foretas en måling etter en viss tid, vil det gi samme resultater som første måling (Walsh & Betz). Deretter bør det finnes samsvar med andre liknende undersøkelser (ekvivalensaspektet ved reliabilitet), samt samsvar dersom testinstrumentet brukes av en annen testleder (vurdererrelabilitet) (Kleven i Lund 2001). Reliabilitet kan variere, avhengig hva slags undersøkelse man gjennomfører.

Testene som er valgt i denne analysen måler ut utsnitt av et barns ferdigheter, og derfor vil reliabilitet være et antatt mål, dersom man testet barnet på nytt med samme test. Det vil alltid være en viss læringseffekt hos barnet som gjør reliabilitet til et vanskelig mål. Men i desto større grad er det viktig å ha med dette statistiske målet, fordi man ønsker at utredningsverktøyet måler det samme hos hvert enkelt barn.

Som nevnt tidligere er det nødvendig å forholde seg til at det finnes et målingsproblem i både psykologi og i pedagogikk, som får betydning for reliabilitet og validitet.

Målingsproblemet dreier seg om at det ofte er ikke observerbare fenomener som ønskes studert. Derfor må man som forsker tillegge det fenomenet som ønskes studert observerbare indikatorer.

Reliabilitetskoeffisient er et statistisk mål, som viser estimerte beregninger av i hvilken grad en test er reliabel altså pålitelig. Det er ulike statistiske metoder for å finne informasjon vedrørende reliabilitet. Reliabilitet er først og fremst relevant for å vurdere om en test, representerer tilfredsstillende operasjonalisering av et begrep (Christophersen 2004). Spørsmål rundt hva en test måler, regnes som irrelevant i reliabilitetsteori, men står mer sentralt innenfor validitetsaspektet (Kleven i Lund 2002).

### **Validitet**

Begrepet validitet viser til i hvilken grad en test måler det den utgir/tilsier at den skal måle. Dersom en test er designet for å måle intelligens, og den i virkeligheten måler noe annet, er testen lite valid som mål av intelligens (Walsh & Betz 1995). Som med reliabilitet, finnes det ulike former for validitet. Derfor må validitetsaspektet ses i sammenheng med hva måleresultatet skal brukes til (Lund 2002). I hvilken grad innholdet i en test er representativt i forhold til det som ønskes målt, kalles innholdsvaliditet (Walsh & Betz 1995).

Kriterievaliditet refererer til sammenhengen mellom en testskåre og en form for kriterie eller resultat. Et eksempel på kriterievaliditet kan være en test som predikerer senere skole eller yrkesprestasjoner (Kleven i Lund 2002). Begrepsvaliditet er kanskje den viktigste formen for validitet, fordi den viser til grad av samsvar mellom begrepet slik det er definert teoretisk og slik begrepet er operasjonalisert (ibid).

### **Operasjonalisering**

Operasjonalisering er som nevnt en vanskelig prosess. Man har et ønske om å måle noe som i realiteten ikke lar seg observere, og det blir derfor også vanskelig å måle. Det er snakk om konstruerte teoretiske begreper, som måles gjennom synlige indikatorer. Dermed blir hvordan man operasjonaliserer begrepet et vesentlig moment, fordi det representerer hvilke synlige variabler som indikerer det abstrakte, konstruerte begrepet (Befring 2002).

Operasjonalisering dreier seg om hvordan man knytter empiriske indikatorer til det teoretiske begrepet som ønskes målt. Dette er problematisk fordi det ikke finnes observerbare indikatorer som dekker et teoretisk begrep fullstendig. Operasjonalisering har betydning for begrepsvaliditet, som viser til samsvaret mellom hvordan man har definert begrepet teoretisk og hvordan begrepet er operasjonalisert (Kleven i Lund 2002).

Når det gjelder operasjonalisering i de testene som er under analyse i denne oppgaven, har det blitt foretatt en rekke operasjonaliseringer i forhold til det som ønskes målt i de forskjellige testene. Dette er en komplisert og tidkrevende prosess. Som vanskelig lar seg beskrive her, fordi det trengs inngående informasjon fra testutviklerne. Derfor vil ikke dette bli omtalt ytterligere.

### **Normering/ Standardisering**

Som nevnt tidligere vil en standardisert test ha en form for normering. En normering foregår ved at et representativt utvalg gjennomfører testen, slik får man mulighet til å se nærmere på resultatfordelingen i dette utvalget. Dette blir sammenligningsgrunnlaget for videre bruk av

---

testen. Det betyr at et barns skåre sammenlignes med normalfordelingen som fremkommer fra utprøvingen fra det representative utvalget (Kleven i Lund 2002). Normalfordeling brukes som et hjelpemiddel i sammenheng med måling av ulike variabler. Når det gjelder menneskelige ferdigheter og egenskaper, er det vanskelig å finne ut hva som representerer den riktige fordelingsformen. Dette er fordi det er mange faktorer som spiller inn i forhold til hvordan resultatfordelingen ser ut. Blant annet vil oppgavens vanskelighetsgrad og utvalgets kunnskaper ha betydning (ibid).

### **Representativitet**

Under standardisering av en test, kreves det at normeringsgruppen er representativ i forhold til målgruppen, det vil si den aktuelle gruppen som ønskes undersøkt. Er normeringsgruppen representativ for populasjonen som ønskes undersøkt? Dersom man har et representativt utvalg, er dette et utvalg som ligner populasjonen så mye at resultatene som fremkommer i utvalget kan regnes som gyldige for populasjonen. Det er viktig å være klar over at det er forskerne som definerer gruppene det trekkes slutninger fra. Det er derfor svært sjelden at man er i stand til å komme frem til et utvalg som representerer hele befolkningen. Det er derfor snakk om representativitet i forhold til utvalgte grupper som det kan trekkes slutninger fra (Kleven i Lund 2002).

Generalisering av resultatene fra en undersøkelse bør ses i samsvar med normeringsutvalget. Man er ute etter å finne ut med hvor stor sikkerhet man kan gå ut i fra at testen viser til et gjennomsnitt for en valgte populasjonen, for eksempel barn i alderen 6 til 16 på østlandsområdet. Minoritetsspråklig bakgrunn er ofte ikke tatt med i standardiseringen. Derfor kan man som regel ikke generalisere resultatene fra en test til denne barnegruppen. Er oppgavene i testen representative i forhold til det teoretiske begrepet som ønskes målt? Dette dreier seg om begrepsvaliditet. Det er altså snakk om i hvilken grad testen måler det som testen tilsier at den skal. Hvorvidt et utvalg er representativt avhenger av hvordan utvelgingen har foregått.

## 5. Empirisk del

I dette kapitlet vil psykologisk testing gjennomgås på et generelt grunnlag. Som et eksempel på en psykologisk test vil WISC III bli beskrevet. WISC III brukes mye i dag av ulike instanser, blant annet Pedagogisk Psykologisk Tjeneste. Dette er en test som har blitt kritisert mye ved bruk på minoritetsspråklige elever. Grunnen til denne kritikken kan forklares gjennom at WISC III inneholder språklige og kulturelle betingelser som kan komme uheldig ut i forhold til denne elevgruppen. Deretter beskrives analysen av testinstrumentene. Grunnlaget for analysen baseres på disse hovedmomentene som vil bli diskutert gjennomgående: hva testen måler, testens kulturelle kontekst og opprinnelse, testens normeringsutvalg, testens original språk, bildene i testen samt ordvalg i testen. Disse spørsmålene er aktuelle av flere årsaker. Hva testen måler er det et relevant spørsmål fordi det viser til testens validitet og reliabilitetsproblematikk. Dersom testen måler noe annet enn det som testen utgir seg for å måle, vil dette kunne svekke validiteten. Hvis testen måler forskjellige ferdigheter og evner fra hvert enkelt barn, vil dette svekke testens reliabilitet. Hvis testen brukes på en annen gruppe enn det testen er tiltenkt vil dette utgjøre eventuelle feilkilder som kan ha betydning skåren til dette barnet.

Testens kulturelle opprinnelse er av særlig interesse, nettopp fordi hva slags kultur som er representert i testen vil kunne påvirke barnets resultater. Dette har også betydning for testens representativitet, det kulturelle og språklige bidraget i en test bør være representativt i forhold til normeringsutvalget. Dersom en test i stor grad gir et bilde av norsk språk og norsk kultur. Vil dette kunne komme uheldig ut dersom denne testen brukes på barn som ikke har tilsvarende språklige ferdigheter som kreves, samtidig som det kulturelle innslaget er ukjent for barnet. Et eksempel på dette, hentet fra WISC III, kan vise nettopp dette. Oppgaven går ut på å svare på en rekke spørsmål, som tar utgangspunkt i at barnet har læringserfaringer fra norsk skolesammenheng. Et eksempel på dette hentet fra WISC III: ”*Hvor langt er det mellom Oslo og Bergen?*” (Wechsler 1999).

Språkbruken i testen kan inneholde kulturelle innslag, som i mindre eller større grad kan være ukjent for barn med enn annen kulturell, språklig og erfaringsmessig bakgrunn. Hvor testen er utviklet har også betydning for hvilken kultur kontekst som gjenspeiles i testen. BPVS II er utviklet i Storbritannia, og man kan på mange måter si at en direkte oversettelse av BPVS vil kunne gi et engelsksspråklig preg på testen. I forhold til vanskelighetsgrad på



---

ordene kan dette ha en betydning, men det er vanskelig å vise til direkte sammenhenger på dette punktet, siden norsk og engelsk er to ulike språk. Høyfrekvente ord på engelsk, trenger ikke nødvendigvis å være høyfrekvente ord på norsk. En oversettelsesprosess vil også kunne sies å overføre feil som finnes i original testen videre til den norske utgaven. Et for stort fokus på selve oversettelsen og at denne skal være korrekt vil kunne føre til at vesentlige psykometriske egenskaper ved testen forsvinner under oversettelsesprosessen, som videre vil ha betydning for testens reliabilitet og validitet.

Videre stilles det spørsmål rundt normeringsutvalget, fordi dette er aktuelt i den grad minoritetsspråklige barn er tatt med i den norske normeringen. Dersom dette er gjort vil det i større grad være naturlig å bruke denne testen på denne elevgruppen, fordi det da vil være et sammenlignings grunnlag. Dersom gruppen ikke er tatt med i normeringen, vil dette kunne svekke argumenteringen for å bruke denne testen på minoritetsspråklige elever.

*”Elever der norsk ikke var det dominerende språket ble også ekskludert, fordi Språk 6 – 16 er konstruert for barn med norsk språkstruktur.” (Ottem og Frost 2005: 22).*

Følgelig er minoritetsspråklige ekskludert fra normeringen av Språk 6 – 16, og bør derfor ikke brukes på denne gruppen. Men i mangel på andre alternativer, er man nødt til å anta at Språk 6 – 16 også brukes på minoritetsspråklige elever, da særlig i PP- tjenesten.

Til tross for at enkelte manualer påpeker at minoritetsspråklige elever ikke er tatt med i normeringen, brukes ofte testen i større eller mindre grad, i mangel av andre passende utredningsinstrumenter.

## 5.1 Psykologisk testing

Psykologisk testing kan ses på som en prosess for å forstå og hjelpe mennesker å takle ulike problemer (Walsh & Betz 1995). Cronbach definerer psykologisk testing slik: *“systematic procedure for observing behavior and describing it with the aid of numeric scales or fixed categories”* (Cronbach 1990:32).

Utviklingen av psykologisk testing har vært og er en prosess hvor man ønsker å utvikle reliable utredningsverktøy som i økende grad nærmer seg det som ønskes undersøkt hos individet (Sergeant & Taylor 2002). Grunnen til at dette nevnes er vissheten om at både

psykologi og pedagogikk arbeider med lite observerbare fenomener. Så hvordan kan disse måles og tallfestes, for så å sammenlignes med andre individers prestasjoner?

Til tross for dette brukes psykologiske tester, fordi det finnes ingen andre metoder som kan komme frem til resultater som i like stor grad kan sies å være et utsnitt av menneskelig atferd. Sett sammen med en lærers uttalelser eller foreldres meninger vil et testresultat kunne gi langt mer objektive beskrivelser av barnet. I tillegg er det mulig å sammenligne et individs resultat med et gjennomsnitt for en utvalgt populasjon.

Oftest vil psykologisk testing inneholde en viss form for systematikk. Disse systematiske prosedyrene forutsetter en rekke ting, blant annet at det forekommer en manual og retningslinjer som følges. Standardiserte tester eller normbaserte tester, som det også kalles, forutsetter at testen er standardisert i forhold til en klart definert gruppe (Sattler 2001).

Normeringsgruppen er et representativt utvalg av den aktuelle befolkningen, som for eksempel norske barns ordforråd i alderen 6 til 16. Det betyr at et barns skåre måles opp i mot normalfordelingen i den aktuelle aldersgruppen.

Standardiseringen refererer til de kontrollerte betingelsene som forekommer under testingen, disse betingelsene skal være konstante. Det vil si at betingelsene med hensyn til administrering i testsituasjon, skåring og tolkning av resultater er de samme hver gang testen gjennomføres. Et eksempel på dette er at instruksjonen under en test, skal være den samme for hvert individ som gjennomfører testen (Walsh & Betz 1995). Dersom instruksjonen er ulik under testingen vil dette kunne gi lite pålitelige resultater. Skåringen av en test bør være så objektiv som mulig. Objektiv skåring av individets prestasjon er en forutsetning for at en test er standardisert.

En testsituasjon vil som oftest være en uvanlig situasjon for mange mennesker. I dagens samfunn er man stadig under en type evaluering, enten i skole eller jobb- sammenheng. En testsituasjon vil likevel oppleves som kunstig og ubehagelig for individet. Til tross for dette er psykologiske testinstrumenter vektlagt i forhold til utredning av vansker av ulik art hos individet. En utredning av vansker hos et barn består som oftest av en innsamling av informasjon for å forstå situasjonen/vanskene til barnet (Sattler 2001). Sattler (2001) deler opp utredning i fire områder: Samtale/intervju, observasjon, uformelle tester og prosedyrer, normerte tester. Denne oppgaven har som utgangspunkt å undersøke utvalgte standardiserte tester.

---

### **Kvantifisering av menneskelige egenskaper.**

Å tallfeste menneskelig atferd eller ferdigheter, er langt mer problematisk enn å måle fenomener i naturvitenskapen. Kvantifisering av menneskelige egenskaper har sine fordeler og ulemper. Kvantifisering av menneskelige egenskaper skaper først og fremst muligheten for å sammenligne et individs ferdigheter i forhold til en bestemt gruppe, for eksempel til barnets jevnaldrende (Sattler 2001). Dette gir mulighet til å undersøke om barnet har en aldersadekvat utvikling på det målte området. Det er for øvrig vesentlig å påpeke at en normbasert test, må ses i sammenheng med annen informasjon rundt barnets situasjon. Både observasjon, samtale med foresatte, og andre uformelle tester gir et helhetsbilde av barnet og dets vansker (Sattler 2001).

Vekt og lengde ses på som stabile måleenheter, som ikke lar seg forandre og påvirke av andre faktorer. Menneskelig atferd er for øvrig mer komplisert å måle. Først og fremst må det fortas en operasjonalisering av variabler av atferd som ønskes studert. En slik prosess forutsetter en ekskludering av enkelte faktorer. Alt kan ikke måles samtidig. Dette bringer oss igjen inn på målingsproblemet som finnes blant annet i pedagogikk og i psykologi. Det som skal måles er ofte et teoretisk begrep, som ikke lar seg observere direkte. Motivasjon er et eksempel på et slikt teoretisk begrep, som er ikke direkte observerbart. Dermed er man nødt til å tillegge observerbare tegn, som representerer begrepet. Som nevnt kalles dette å definere begrepet operasjonelt (Kleven 2002). Det betyr at egenskapene, fenomenene eller faktorene blir operasjonaliserte variabler av et utsnitt menneskelig atferd. I utgangspunktet vil disse variablene alltid være feil, men for å kunne foreta målinger i pedagogikk er man nødt til å tillegge faktorene slike variabler, til tross for usikkerheten rundt. Hvor godt samsvaret er mellom hvordan begrepet slik det er definert teoretisk og hvordan begrepet er operasjonalisert kalles begrepsvaliditet (Kleven 2002).

### **Teste barn med minoritetsspråklig bakgrunn**

Å teste barn kan være en utfordring. Men hva med de barn som har en annen språklig og kulturell bakgrunn? Hvilke retningslinjer følges? Og i hvor stor grad tas det hensyn til barnets kulturelle, språklige og erfaringsmessige bakgrunn? Generelle evnetester som Wechsler Intelligence Scale for Children – III (WISC III) har fått mye kritikk for at den brukes på minoritetsspråklige barn. Wechsler Intelligence Scale for Children er kanskje den mest omtalte, utprøvde og brukte psykologiske testen som måler intelligens, kunnskapsnivå og skoleprestasjoner hos barn (Sattler 2001, Pihl 2005).

Til tross for dette kan det diskuteres hvorvidt WISC III bør brukes på minoritetsspråklige elever, fordi denne gruppen kan oppnå en IQ skåre som er lite pålitelig, på grunnlag av at testen ikke er utviklet med tanke på denne elevgruppen. Enkelte deltester i WISC III måler tilegnet skolefaglig kunnskap. Dermed kan noen minoritetsspråklige barn ha vansker med enkelte av deltestene i WISC III; fordi de har en annen kultur, språk og erfaringsbakgrunn. WISC III stiller samtidig krav til språklige ferdigheter, som igjen kan føre til en dårligere og lite pålitelig skåre hos enkelte minoritetsspråklige elever, fordi noen i denne gruppen kan ha svakere ferdigheter i norsk enn majoritetsspråklige elever. Pihl har i sine undersøkelser kommet frem til at det har blitt foretatt uheldige konklusjoner fra Pedagogisk Psykologisk Tjeneste, som har ført til en kunstig konstruksjon av minoritetsspråklige elever som psykisk utviklingshemmet, basert på testresultater fra en test som ikke er tiltenkt denne elevgruppen (Pihl 2005). Dette er urovekkende i flerkulturelle Norge, og viser at det trengs ulike tester, som kan brukes på denne elevgruppen. Hva slags testinstrumenter skal brukes på minoritetsspråklige elever dersom det trengs? Det er begrenset utvalg, og som nevnt ovenfor er det vanskelig å unngå kritikk gjennom å bruke testene som foreligger. Man kan stille seg kritisk blant annet fordi minoritetsspråklige ikke er tatt med under normering og standardiseringsprosessen av testen. Barnets prestasjoner kan risikere å bli sammenlignet på feil grunnlag. Mangelfull skolegang, ulik kultur, språk og erfaringsbakgrunn utgjør ofte forskjellene mellom minoritet – og majoritetsspråklige elever. Det etterspørres tester til bruk på minoritetsspråklige elever. Men hvordan kan man utvikle en test som kan favne de ulike språkgruppene med variert kultur og erfaringsbakgrunn. Det er lite sannsynlig å anta at en elev med norsk- pakistansk bakgrunn har samme kulturelle, språklige og erfaringsbakgrunn som en elev med norsk- tamilsk bakgrunn. Det kan sies at ingen test er kulturfri, som igjen gjør det ytterligere problematisk å utvikle den ”perfekte” testen.

Dersom barnet har faglige vansker på skolen, bør det foretas en kartlegging, for å avdekke barnets vansker. Minoritetsspråklige elever kan ofte ha et velutviklet hverdagspråk, som skjuler de akademiske vanskene. I noen tilfeller blir disse vanskene mer fremtredende utover i opplæringen. Cummins viser til at minoritetsspråklige ofte har et relativt velutviklet hverdagspråk, mens de har et mer begrenset begrepsapparat i forhold til skolefaglig/akademiske begreper (Cummins 2000). Dette viser hvor viktig det er å foreta en adekvat kartlegging. Dersom det tas i bruk tester som ikke er tiltenkt minoritetsspråklige elever, må dette legges vekt på i tolkningen av barnets resultater (Sattler 2001). Slik unngås feildiagnostisering. De testene som er tatt med i analysen i denne oppgaven, er tester som

---

ikke nødvendigvis har tatt med barns ulike språklige og kulturelle bakgrunn, som da igjen vil kunne ses på som kritikkverdig.

### **Kontroverser rundt bruk av standardiserte tester**

Bruk av standardiserte tester har blitt utsatt for en mengde kritikk gjennom tidene. Kritikken består blant annet av at slike tester ikke kan brukes på ulike etniske minoritetsgrupper, fordi disse gruppene kan ha en etnisk, kulturell, språklig, sosioøkonomisk og erfaringsbakgrunn, som er differensielt forskjellig fra testens normeringsutvalg (Sattler 2001). Videre påstås det at testingen ikke alltid er til barnets beste. Dette kan føre til:

- a) Feilplassering av elever i forhold til spesial undervisning.
- b) At elevene ikke kartlegges riktig slik at deres reelle vansker avdekkes, dermed kan enkelte gå glipp av nødvendig spesialpedagogisk hjelp.
- c) At eleven ikke kartlegges i det hele tatt, som igjen fører til at eleven kan miste spesialpedagogisk tilbud, og tilpasset opplæring innenfor elevens nivå.

Ofte foregår testingen på et språk som elevene ikke behersker fullt ut, dette kan få betydning for elevens resultat. Samtidig kritiseres det fra en rekke hold at testingen ofte gjennomføres uten fullstendig samtykke fra foreldre. Pihl (2005) viser til at Pedagogisk psykologisk tjeneste har brukt termer og uttrykk som er lite forståelig for foresatte. Å bruke faglige termer i forhold til barnets vansker, kan misforsås av foresatte. Fordi de ikke er klar over hva som ligger i de ulike faglige termene som brukes for å forklare barnets vansker. Personer uten pedagogisk eller psykologisk bakgrunn vil ikke ha noen forutsetning for å forstå hva som ligger i for eksempel betegnelsen: Generelle lærevansker. Som betyr at barnet har evnemessige ferdigheter som er langt under gjennomsnittet. Pihl (2005) viser til at det foregår en diskurs i visse faglige miljøer, som fører til at foresatte, lærer og andre føres bak lyset, når det brukes faglige termer som forklaring på barnets vansker. Man kan si at en diskurs har en rekke regler for hvem som skal kunne uttale seg, og hvilke hensyn som blir tatt i forhold til dette (ibid).

Standardiserte tester kan sies å gi et begrenset mål på hva som egentlig er barnets evner og ferdigheter (Green 1981).

*”A standardized test is a task or a set of tasks given under standard conditions and designed to assess some aspect of a person’s knowledge, skill or personality” (Green 1981).*

Nettopp fordi testen i seg selv bare er et mål på et utsnitt av menneskelig atferd. Til tross for omfattende kritikk brukes standardiserte tester i stort omfang. Testene kan fortelle noe om barnets evner og ferdigheter, sammenlignet med et gjennomsnitt i populasjonen. Dersom testene administreres riktig og tolkningen tar hensyn til etiske og kulturelle forskjeller, skal de i følge Sattler (2001), kunne gi valide resultater.

*”Test advocates believe that standardized tests, if they are administered and interpreted carefully and ethically, have valid uses” (Sattler 2001:27).*

### **Tolkning av resultater**

En tolkningsprosess består ikke kun av en tolkning av barnets testskåre. Ved å sette fokus på ett tall, står man som utreder i fare for å overse vesentlige momenter i utredningen. Gjennom å tolke testprosessen både kvalitativt og kvantitativt vil en få mulighet til å se barnets begrensninger, samt dets sterke sider, i en og samme prosess (Sattler 2001).

*“Each child has a range of competencies and limitations that you can evaluate by both quantitative and qualitative means... the focus should not be only on problems.” (Sattler 2001:9).*

Det er imidlertid viktig å være løsningsfokusert, men fokuset bør ikke ene og alene ligge på barnets problemer (ibid). Ulike faktorer vil kunne påvirke en testskåre. Derfor bør den som tolker testresultatene til barnet, se resultatene i lys av en rekke andre faktorer som kan være av betydning. For det første har barnet medfødte evner av både spesifikk og generell art. I tillegg har barnet bakgrunnsfaktorer av kulturell, språklig, erfaringsmessig og etnisk art. Dersom barnet har vært borte i liknede oppgaver, kan det ha bedre forutsetning til å løse oppgavene, enn barn som ikke har løst slike oppgaver. For øvrig vil også personlighetstrekk som motivasjon, angst og interesse være av betydning for testskåren. Deretter kan situasjoner som oppstår i under selve testingen, instruksjon fra testleder, barnets oppfattelse av testen, miljømessige påvirkninger som støy og liknende, innvirke på barnets testskåre. Videre vil kravene som ligger i selve testen ha betydning for skåren, enten det er snakk om krav til spesifikke ferdigheter eller krav om tid for å få poeng på oppgaven. Til slutt vil tilfeldig variasjon, som for eksempel gjetting og skrivefeil påvirke testresultatet (Sattler 2001).

---

## 5.2 WISC III - et verktøy for kartlegging av intellektuell kapasitet hos barn.

WISC III (Wechsler Intelligence Scale for Children- Third Edition) en av de mest brukte, omtalte, anvendelige og pålitelige psykologiske testen som måler intelligens, kunnskapsnivå og skoleprestasjoner hos barn (Wechsler 1999, Sattler 2001, Pihl 2005).

*”Wechsler Intelligence Scale for Children- Third Edition (WISC III) är ett psykologiskt mätinstrument som administreras individuellt med syfte att bedöma den intellektuella kapaciteten hos barn i åldrarna från 6 år upp till och med 16 år och 11 månader”.*  
(Wechsler 1999: 9).

WISC III er et måleinstrument som administreres individuelt med formål å bedømme kognitive funksjoner hos barn i alderen fra 6 år og opp til 16 år. WISC består av ulike deltester som måler ulike aspekter av intelligens. Dette sammenfattes i tre ulike poengsummer: Verbal IQ, utførings IQ, og total IQ. I tillegg har man mulighet til å se på fire faktorbaserte indeks-poeng (Wechsler 1999). Standardiseringen av WISC III er foretatt i Sverige, og det er disse normene som brukes i Norge. Det har vist seg at normeringsutvalget fra Sverige, består av barn fra et område i Sverige hvor foreldrene har høyere utdanning, enn resten av gjennomsnittspopulasjonen i Sverige. Dette har blitt forsøkt rettet opp, men kan ses på som en svakhet ved testens normeringsutvalg. (Wechsler 1999). Minoritetsspråklige elever er ikke tatt med i normeringen, og det påpekes i manualen at lave resultater på WISC III kan være et resultat av ekstrem kulturell og/eller språklig forskjeller.

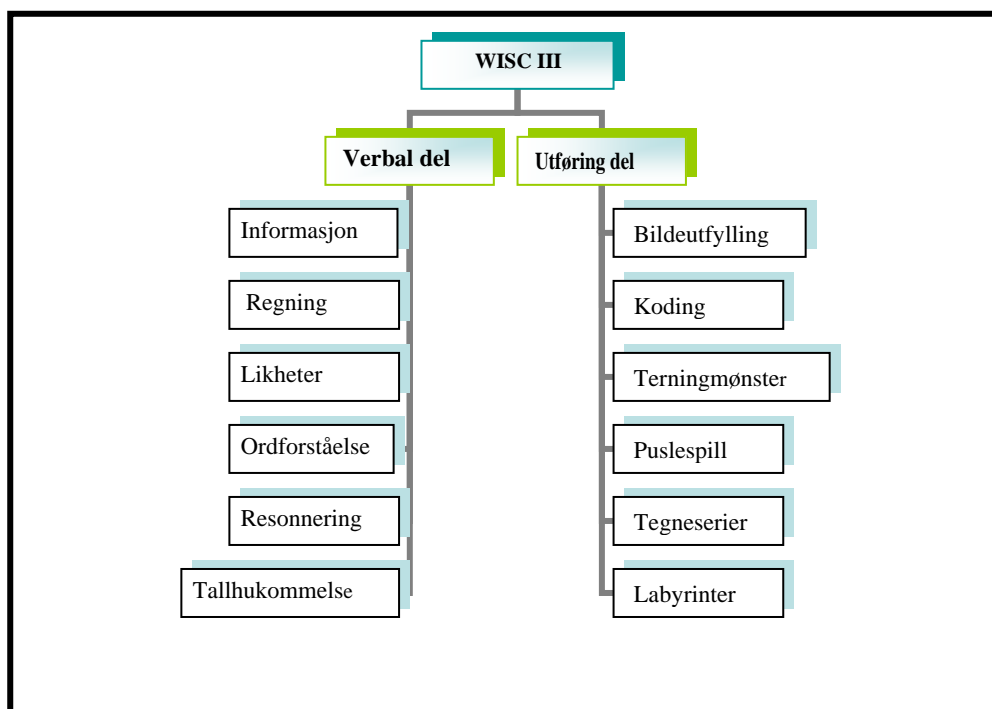
*”Om barnet kommer från en annan kultur och/eller har bristfälliga kunnskaper i svenska, kan testen knappast fungera så som det er avsett”* (Weshsler 1999:89).

Dersom WISC III brukes på barn som kommer fra en annen kultur, eller har dårlige ferdigheter i svensk, kan dette gå ut over barnets resultater og dermed vil ikke testen kunne fungere som tiltenkt, nemlig å måle intellektuell kapasitet (Weshsler 1999). En kan derfor anta at dette også gjelder for den norske utgaven av WISC III, som nevnt bruker svensk normering.

WISC III er en statistisk verbal evnetest, som kan brukes innenfor aldersspekteret 6 til 16 år. Den engelske utgaven er ikke egnet for de som ikke behersker engelsk. Følgelig er testen beregnet for de som kan landets majoritetsspråk. WISC III er normert for svenske barn, og inneholder som man kan se av den skjematiske oversikten over 13 deltester. Testen har stoppregel på tre feil, som betyr at man avbryter en deltest dersom barnet får tre feil på rad. Enkelte deltester har også en tidsbegrensning, som betyr at dersom denne overskrides får barnet 0 poeng. Det er fare for læringseffekt ved re- testing, og man bør undersøke nøye om barnet har blitt testet med WISC III tidligere, og hvor lenge siden denne testingen ble gjennomført. WISC III bør bare brukes av de som har kunnskap om testen og psykologiske forhold rundt dette. WISC III oppfyller psykometriske krav.

### FIGUR 1: WISC – III

Nedenfor vises en skjematisk oversikt over de ulike deltestene i WISC III.



**Informasjon** – En serie spørsmål, hvor barnet skal svare på en rekke spørsmål, som dreier seg om hva slags kunnskaper barnet sitter inne med i forhold til vanlige hendelser, formål, plasser og mennesker. Barnet svarer muntlig.



---

**Bildeutfylling** – fargebilder med vanlige hendelser og personer, ting, eller dyr, hvor en vesentlig del mangler, barnet skal svare eller peke ut riktig del.

**Koding** – En rekke symboler som barnet skal kopiere og pare sammen med geometriske figurer eller siffer. Barnets alder avgjør hvilket kodings ark som skal brukes.

**Likheter** – En rekke spørsmål som barnet skal svare og forklare likheter mellom ulike begreper.

**Tegneserier** – Ulike fargede bilder som presenteres i feil rekkefølge, barnets oppgave blir å sette de sammen slik at historien på bildene gir mening.

**Regning** – Serier med ulike regneproblem som barnet skal løse muntlig.

**Terningmønstre** – Todimensjonelle geometriske mønstre presentert på et ark, barnets oppgave blir å avbilde et bilde, ved hjelp av tofargede klosser

**Ordforståelse** – En rekke spørsmål, hvor barnet skal definere ord muntlig.

**Puslespill** – Puslespill av bilder med vanlige objekter, som barnet skal forsøke å legge sammen.

**Resonnering** – Spørsmål som krever at barnet kan svare hvordan man kan løse hverdagslige problemer, samt forstå sosiale regler og begrep.

**Tallhukommelse** – Muntlig presenterte tallserier hvor barnet skal gjenta, både forlengs og baklengs.

**Labyrinter** – En rekke labyrinter, som barnet skal løse ved hjelp av penn. Barnet skal tegne en strek fra midten, og finne veien ut av labyrinten uten å gå feil vei, og ikke gjennom noen ”vegger”.

**Symbolleting** – En serie iblandet grupper med ulike figurer, barnet skal forsøke å finne igjen utvalgte figurer, som fins i gruppen (Wechsler 1999).

### 5.2.1 Hva måler testen

WISC III er utviklet for å bedømme intellektuell kapasitet hos barn i alderen 6 til 16 år.

Barnet kommer ut med en IQ indeks og fire faktorbaserte indekspoeng. WISC III kan brukes til ulike formål, blant annet psykologisk bedømming i forhold til utdanning og undervisning, pedagogiske støtteinstanser, identifisering av ulike kognitive profiler i forhold

til innlæringsvansker, klinisk og nevropsykologisk bedømming samt forskning (Wechsler 1999).

### **5.2.2 Testens kulturelle kontekst og opprinnelse**

Testen er opprinnelig utviklet i USA. Men har blitt oversatt til en rekke andre språk. I Norge brukes den svenske utgaven, som er oversatt og normert i Sverige (Wechsler 1999).

### **5.2.3 Testens normeringsutvalg**

WISC III er standardisert på 2200 barn i USA. Denne standardiseringsgruppen er et representativt utvalg av barn i USA. Den svenske normeringen har tatt utgangspunkt i den britiske versjonen av WISC, da den antas å ligge nærmere svenske forhold enn den amerikanske versjonen. Den verbale delen inneholder mer kulturelle betingelser og er derfor tilpasset det språkområdet testen skal brukes i (Wechsler 1999). Normene baseres på et standardiseringsutvalg som representativt for den svenske barnepopulasjonen. Utvalget er hentet fra førskoler og skoler i området rundt Uppsala og Stockholm fra barn i alderen 6 til om med 15 år. Utvalget består av 1036 barn. Kravet for å delta i normeringsutvalget var at barna skulle klare å delta i vanlig førskole eller skole. I ettertid har det som sagt, vist seg at det svenske utvalget, har foreldre som har ett høyere utdanningsnivå enn resten av Sverige (Wechsler 1999).

Pedagogisk psykologisk tjeneste bruker ofte WISC III i sine utredninger av barn og unge. PP-tjenesten har i mange tilfeller blitt kritisert for å bruke denne testen på språklige minoriteter (Pihl 2005). Årsaken til kritikken ligger i at WISC III inneholder både kulturelle og språklige betingelser, som kan føre til at minoritetsspråklige elever får en lav IQ skåre. Dette kan føre til feilplassering av disse elevene i forhold til spesial undervisning, når enkelte av disse elevene kanskje har større behov for morsmålsopplæring og tospråklig fagopplæring. Det er særlig de verbale oppgavene i WISC III, som kan være en utfordring for språklige minoriteter. Det har vist seg at denne elevgruppen gjør det dårligere enn majoritetselevne på verbale oppgaver (Cummins 1984, Pihl 2005). Forutsetningene for testens validitet svekkes ved å bruke WISC III på språklige minoriteter, særlig med tanke på at de er utelatt i normeringen.

---

I henhold til Cummins, er prestasjonene som testene måler et produkt av et avansert samspill mellom pedagogiske, kulturelle, relasjonelle og institusjonelle variabler i skole og samfunn som påvirker barnets prestasjoner (1984). Dermed ser man tydelig at enkelte tester preges av språklige og kulturelle forhold som er avgjørende for barnets skåre. Det er gjort undersøkelser rundt språklige minoriteter og bruk av WISC R, som er forgjengeren til WISC III, hvor man har kommet frem til at denne elevgruppen gjør det vesentlig bedre på utføringsdelen i WISC R, enn den verbale delen. Dette kommer ikke som en overraskelse, da den verbal delen i større grad er et utsnitt av elevens skolemessige, språklige og kulturelle forutsetninger. Det er særlig deltesten *likheter*, som viser seg å reflektere hvite engelskspråklige, middelklasse verdier. Denne deltesten ekskluderer læringserfaringer som er gjort i andre kulturelle og språklige miljø (Cummins 1984). Det er viktig som fagkyndig å stille spørsmål rundt bruken av IQ tester på minoritetsspråklige elever, og i hvilken grad disse bør brukes. Det er også nødvendig å fokusere på når slike tester vil kunne gi valide resultater for denne elevgruppen. Hvor lenge bør disse elevene ha vært i de respektive landene for at slike tester kan gi gyldige og pålitelige resultater? Cummins har vist at det kan ta opp til fem år før denne elevgruppen har tatt igjen majoritetsspråklige elever på den akademiske fronten (Cummins 2000). Dette kan bety at denne elevgruppen kan ha vansker på skolen i forhold til faglig utvikling, mens de fungerer sosialt og kommunikasjonsmessig. Intelligenstesting ses på som et ferdighetsmål, og følgelig vil det innebære et mål av tidligere læringserfaringer, som igjen vil være påvirket av kultur og språk. Dermed vil IQ- testing være et mål på dette. Elever som ikke har samme læringserfaringer som gjennomsnittspopulasjonen vil kunne få problemer under en slik eventuell testsituasjon.

Som diskutert ovenfor er det en rekke kritikkverdige momenter ved intelligenstesting av språklige minoriteter. Oppsummert kan man si at det er vesensforskjeller mellom minoritetsspråklige og majoritetsspråklige som kan utgjøre forskjeller i barnets resultat. Ofte er det både språklige, kulturelle og erfaringsmessige barrierer som kan komme frem i en testsituasjon, og påvirke barnets resultat, slik at det blir lite pålitelig. Dermed ser man at det finnes et behov for utvikling av tester som er mindre kulturelt og språklig betinget. Slik vil man kunne gjennomføre en adekvat kartlegging av minoritetsspråklige elevers norsk språklige ferdigheter. I forhold til det kartleggingsverktøyet som foreligger per dags dato, er det få verktøy som er designet spesifikt for kartlegging av språk hos minoritetsspråklige barn. Leiter R er et alternativ, som er en non- verbal intelligens test. Denne testen er mindre kulturell og språklig påvirket enn WISC III. Men selv Leiter R inneholder kulturelle

betingelser, som kan føre til et feilaktig resultat i forhold til den nevnte elevgruppen.

Cummins viser til at begrepsvaliditeten i en IQ test forsvinner dersom testen brukes på barn som har læringserfaringer som er differensielt forskjellig fra den gruppen testen er standardisert for (1984).

## 6. Språk 6 -16

Språk 6 -16 er utviklet av Jørgen Frost og Ernst Ottem fra Bredtvet kompetansesenter (2005). Testen har til hensikt å identifisere barn med språkvansker. Testen kan være et godt supplement for utfyllende informasjon vedrørende barnets språkvansker eller som utgangspunkt for eventuell viderehenvisning til andre instanser som for eksempel Pedagogisk Psykologisk tjeneste. Testen viser hvorvidt barnets språkutvikling er aldersadekvat. Testen bygger på antakelser om at det finnes en sammenheng mellom *begrensninger i verbalt arbeidsminne og utvikling av vokabular* (Ottem og Frost 2005). Barn med språkvansker har ofte mer en ett problem, dette kan gjøre utredningen av slike vansker ytterlig problematisk (ibid). Testen består av en screening del som inneholder fire deltester, samt åtte supplerende deltester, som belyser barnets ferdigheter innenfor *fonologisk bevissthet, grammatikk, og lesing* (Ottem & Frost 2005). Screening delen fra Språk 6- 16 er tatt med i analysen, dette valget er tatt fordi de andre supplerende deltestene i større grad berører barnets leseferdigheter og fonologiske utvikling. Disse ferdighetene har en sterk sammenheng med språkutvikling, men i dette tilfellet var det vesentlig å velge kun et fokusområde. Dette gjøres på grunn av oppgavens problemstilling, og fordi det ikke stilles noe krav til leseferdigheter i de andre testene som analyseres. De supplerende deltestene omtales ikke nærmere. Men det bør imidlertid nevnes at det viser seg at en del barn med språkvansker, står i større risiko for også å utvikle lese- og skrive vansker (Hoskyn 2004).

### 6.1.1 Hva måler testen

Screeningdelen i Språk 6 – 16 er delt opp i fire deltester:

- 1) *Setningsminne,*
- 2) *Ordspenn,*
- 3) *Motsetninger og*
- 4) *Ordkunnskap.*

Deltestene viser til relativt ulike elementer av språket. Deltestene viser til mål av underliggende separate kognitive ferdigheter, dette betyr i henhold til Ottem at de ulike deltestene måler ferdigheter som kan være relativt uavhengig hverandre (Ottem i Skolepsykologi 2007).

*”Det vil si at resultatene på en deltest er relativt uavhengig av resultatet på de to andre fordi de tre deltestene kartlegger separate kognitive funksjoner som hver på sin måte bidrar til utviklingen av språkrelaterte vansker” (Ottem i skolepsykologi 2007:26).*

Ordspenn måler det språklige korttidsminnet. Setningsminne antas å gi et mål på det språklige minnet og individets begrepsutvikling. Motsetninger og Ordkunnskap gir et mål på språklige begreper. Den underliggende antakelsen bak Screeningdelen i Språk 6 – 16, viser til en sammenheng mellom verbalt arbeidsminne og barnets vokabularutvikling (Ottem og Frost 2005). Det er særlig den fonologiske løkken som sies å ha betydning for innlæring av språk. Den fonologiske løkken antas og være en del av arbeidsminnet, og har betydning for innlæringen av nye ord. Dersom individer har en svikt i den fonologiske løkken vil dette kunne produsere språkvansker av ulik art hos individet. Den fonologiske løkken er spesialisert for å holde på verbal informasjon over en kort tid. En svikt her vil kunne føre til innlæringsproblemer i forhold til språkutvikling, og dermed medføre språkrelaterte vansker (Baddeley, Gathercole mfl 1998). Den fonologiske løkken skal hovedsakelig hjelpe til slik at individet lærer nye ord (ibid). Gjennom mål av verbalt korttidsminne, viser det seg at barn som har et sterkt verbalt korttidsminne, i tillegg har et velutviklet vokabular (Ottem, Lian & Karlsen 2007).

Deltesten *Setningsminne*, utføres ved at barnet får opplest en setning fra testleder, som barnet skal gjenta muntlig. Barnet må gjenta setningen korrekt for å oppnå full skåre. Setningslengden øker relativt hurtig, og varierer i antall fra 4 til 18 ord. Dette skal gi et anslag på hvorvidt barnet har vansker med å fastholde og organisere strukturen i setningen (Ottem and Frost 2005). I deltesten *Ordspenn* får barnet opplest ulike ord etter hverandre, som skal gjentas av barnet. Denne deltesten er delt opp i tre blokker, hvor hver blokk inneholder fire oppgaver, med ulike ord som barnet skal gjenta. Oppgavene starter med at barnet skal gjenta tre ord, fire ord, og deretter fem ord. Barnet får opplest: ”Tann – Sekk – Kokk”, og skal deretter gjenta. Ordramsene rimer annenhver gang (ibid). *Ordspenn* måler hvorvidt barnet har problemer med å fastholde lydstrukturen i ord. I deltesten *Motsetninger* skal barnet finne motsetninger av det ordet som leses opp av testleder. Barnet får til å begynne med presentert hele setninger, som for eksempel: ”Er kapteinen liten eller er han -- ?” Senere stilles barnet spørsmål, som for eksempel: ”Hva er det motsatte av å Starte?”.

Deltesten *Ordkunnskap* består av en rekke spørsmål om ords betydninger, som for eksempel: ”Hva er en hatt?” Begge disse deltestene måler barnets evne til å forstå og uttrykke ulike ords betydning (Ottem og Frost 2005). Ordkunnskap kan sammenlignes med deltesten *Ordforståelse* fra WISC III, som har relativt likt utgangspunkt som Ordkunnskap. Det dreier seg om å definere og forklare ulike ords betydning. Flere validitetsaspekt har blitt undersøkt i Språk 6 – 16, blant annet kriterievaliditet, som viser til samsvar mellom testskåre og en eller flere kriterier. Her ble lærerne bedt om å vurdere elevens språklige ferdigheter, sett opp i mot barnets testskåre. Det var høyt samsvar mellom lærers vurdering og barnets testskåre, og man kan følgelig gå ut i fra at Språk 6 – 16 har en *god kriterievaliditet* (Ottem & Frost 2005). *Innholdsvaliditeten* er vurdert å være god, gjennom å bruke Språk 6 – 16 på en gruppe elever som man antok hadde språkvansker, fant man at disse barnas råskåre stemte overens med antakelsen man hadde fra før av (ibid).

Indre konsistens brukes som et mål på reliabilitet. Det innebærer at man undersøker enkelt items i det området som testes, og i hvilken grad disse antas å måle en underliggende dimensjon. Man er altså interessert i hvorvidt ulike deler av testen måler samme sak. Språk 6 – 16 reliabilitets koeffisienter er beregnet med bruk av Cronbachs Alfa. Screening delen i Språk 6 – 16 har en reliabilitetskoeffisient på 0,93 (Ottem & Frost 2005).

### **6.1.2 Testens kulturelle kontekst og opprinnelse**

Hvilken kulturelle kontekst kan sies og ligge bak Språk 6 – 16? Språk 6 – 16 er som nevnt utviklet fra Bredtvet kompetanse senter. Testen har bakgrunn i norsk kultur og kan derfor sies å være en god test for å avdekke språkvansker hos norske barn. Men hva med de barna som ikke har samme kulturelle, språklige og erfaringsmessige bakgrunn som norske barn. Stiller disse barna svakere med tanke på Språk 6 – 16? Det vil alltid være kulturelle variasjoner i forhold til hvilken kultur et barn vokser opp i. Men det som er avgjørende her er om minoritetsspråklige barn vokser opp i en kultur, eller har en kultur, et språk og en erfaringsbakgrunn som avviker fra det som representeres i Språk 6 – 16. Dette gjør det mer problematiske å bruke Språk 6 – 16 på språklige minoriteter. Manualen påpeker at testen bare skal brukes på barn som har norsk som førstespråk (Ottem & Frost 2005).

Det er en rekke punkter som bør nevnes i forbindelse med testens kulturelle kontekst og opprinnelse. For det første inneholder en av deltestene rim. Det antas at norske barn i større grad er kjent med rim og regler enn språklige minoriteter. Rim og regler presenteres for

norske barn relativt tidlig barnsdommen, ofte i forbindelse med høytlesning hjemme, og deretter i barnehagen. En av deltestene i Språk 6 – 16 innebærer å holde på et visst antall ord, hvor enkelte av disse ordene rimer. Det kan virke mer sannsynlig at barn som er kjent med rim, vil kunne mestre denne oppgaven bedre enn barn som ikke har kjennskap eller erfaring til rim. Valg og ord og begreper har også en del å si for testens kulturelle kontekst. Det er spesielle ord og begreper som er langt mer kjent for norske barn, enn for språklige minoritetsbarn. Det er særlig i deltesten *Setningsminne* hvor barnet blir bedt om å gjenta setninger i stadig lengre versjoner. Her er det er del setninger og situasjoner som for mange minoritetsspråklige vil være ukjente, dette vil dermed kunne gi utslag på barnets resultater. Å huske ord/setninger som en kan relatere til seg selv og egen virkelighet vil være lettere å gjenta, enn setninger som ikke gjør det. Språk 6 – 16 kan sies å være lite representativt i forhold til minoritetsspråklige elevers kulturelle kontekst og opprinnelse, da Språk 6 – 16 i stor grad representerer norsk middelklasse kultur.

### **6.1.3 Testens normeringsutvalg**

Normeringsutvalget i Språk 6 – 16 var opprinnelig 1080 personer innenfor aldersspekteret 6 til 16 år. Fordelingen av utvalget kommer fra fire ulike geografiske områder i Norge (Ottem og Frost 2005). Normeringsutvalget ble ved en senere anledning økt til 1214 barn innen for det nevnte aldersspekter. Dette ble gjort da det ble etterspurt en mer fingradert skala som lettere lar seg sammenligne med andre testresultater (Ottem i Skolepsykologi 2007). Barn med annet morsmål enn norsk ble ekskludert fra normeringen, siden Språk 6 – 16 er konstruert for norsk språkstruktur (Ottem og Frost 2005). Dermed betyr dette at man ikke bør bruke Språk 6 – 16 for å avdekke språkvansker hos minoritetsspråklige. Dersom dette gjøres, vil det være mindre sannsynlig at barnets resultater kan være valide og reliable. Fagpersoner bør derfor være ytterst forsiktig med hva som tillegges i en tolkning av minoritetsspråklige barns testresultater fra Språk 6 til 16.

For å avdekke språkvansker hos minoritetsspråklige elever er det nødvendig at det utvikles en parallell og bedre tilpasset test på barnets minoritetsmorsmål. Øzerk (2003) har i sine undersøkelser kommet frem til at det er 14 ulike morsmål, som dekker opptil 78 % av andelen minoritetsspråklige elever i Oslo skolen. Dersom det ble utviklet en parallell test på disse 14 ulike språkene, ville man være bedre rustet til å avdekke språkvansker også hos denne elevgruppen.



#### 6.1.4 Ordvalg i testen

Det har blitt utført en analyse av de ulike ordene i Språk 6 – 16 for å vise de ordvalgene som er foretatt under utviklingen av testen. Som man kan se av figur 2, er det en vesentlig større andel av substantiv og verb i Språk 6 – 16.

Det er verdt å merke seg at i deltesten *Setningsminne* øker antall ord nokså hurtig etter oppgave nr 4. Dette betyr at deltestens vanskegrad øker kanskje fortere enn nødvendig. Ordvalgene i denne deltesten er også av betydning, da enkelte av ordene kan ses på som nokså utdaterte og lite aktuelle. Minoritetsspråklige elever vil på mange måter ikke kjenne seg igjen i situasjonen som beskrives, det kan også hende at barnet aldri har hørt ordet før. Dette gjør det vanskeligere for barnet å kunne gjenta setningen. Når enkelte av ordene i oppgaven fremstår som nonord for barnet. Det er kjent at individet lettere holder på informasjon som er kjent, fremfor ukjent informasjon (Helstrup & Kaufmann 1999). Dette er fordi individet da har en tidligere representasjon av ordet som fører til at arbeidsminnet holder lettere på informasjonen, slik at barnet har mulighet til å gjengi oppgaven korrekt. Et eksempel på dette kan være: ”*Mor og far reiste ut på tur i en liten kano.*” Begrepet ”tur” kan betegnes som et særnorskt middelklasse fenomen. Spørsmålet er hvilke familier som drar på kanotur, og hvor mange barn vil kunne kjenne seg igjen i denne situasjonen? En kan anta at det er stadig færre barn som har erfaring med å dra på kanotur, og de som vil kjenne seg igjen i en slik situasjon, mest sannsynlig vil tilhøre middelklassen, med norsk som morsmål. Dermed vil mange minoritetsspråklige barn har mindre kjennskap til det å dra på kanotur, som kan være lite fordelaktig i forhold til barnets resultater på Språk 6 – 16. I tillegg kan ordet ”kano” være ord som mange barn ikke har hørt før, og da spesielt minoritetsspråklige elever, men også norske majoritetselever. Dette byr på ekstra vansker for å kunne holde og gjenta denne spesifikke setningen.

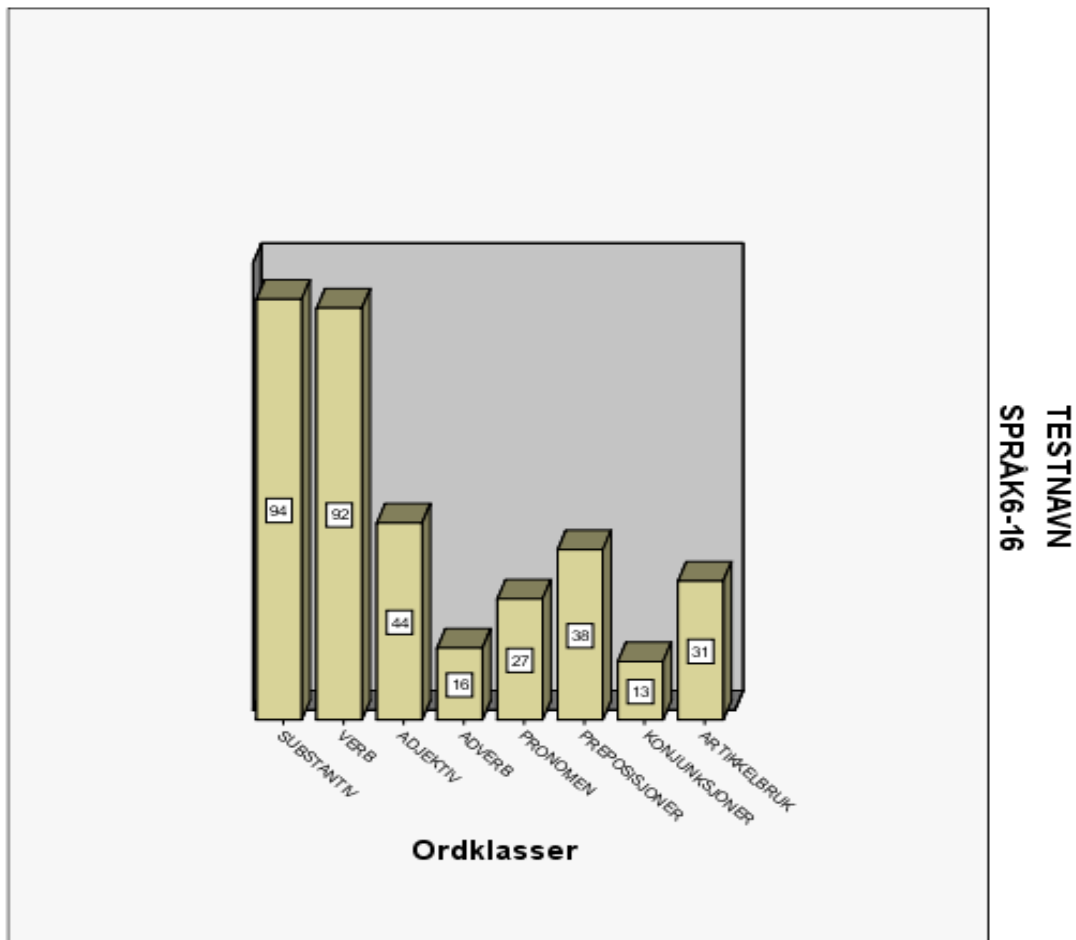
Når det gjelder ordtypene som er tatt med en Språk 6 – 16, er det nokså variert utvalg av ordtyper. Det er flest substantiv, påfulgt av verb. Noen av ordene kan ses på som nevnt som lite aktuelle i forhold hva slags ord dagens barn har ervervet. Dette gjelder spesielt ”*fure*” og ”*fabel*” i deltest 4.

Dersom barnet er vant med norske læringserfaringer vil Språk 6 – 16 være et tilfredsstillende kartleggingsverktøy. Men hvis barnet har læringserfaringer fra en annen kulturell, språklig og erfaringsbakgrunn vil Språk 6 – 16 være mindre hensiktsmessig med tanke på språklig

kartlegging, og spesielt enda mer problematisk dersom testen gjennomføres på barnet svakeste språk.

Figur 2 nedenfor viser en oversikt over ulike ordtyper i Språk 6 – 16. Av tabellen fremkommer det at Språk 6 – 16 inneholder flest substantiv, påfulgt av verb, adjektiv, preposisjoner, pronomener, artikler, adverb og til slutt konjunksjoner.

**Figur 2: Ordklasser i Språk 6 – 16**



Det forekommer ingen bilder i Språk 6 – 16, og det blir derfor ikke omtalt nærmere.

## 7. Test for Reception of Grammar 2ed – Trog 2

TROG 2 (Test for Reception of Grammar) er utviklet av Dorothy Bishop. I utgangspunktet var TROG et forskningsverktøy, som opprinnelig skulle brukes for å avdekke visse aspekt ved *språkforståelsen hos barn med spesifikke språkvansker* (Bishop 1989; Lyster 2007).

Testen er under oversettelse og normering av Spesialpedagogisk institutt (ISP), Universitetet i Oslo. TROG er i følge den norske forløpige manualen oversatt og tilpasset norske forhold. *Testen kartlegger forståelse av grammatiske kontraster gjennom bøyninger, ord funksjoner og ordstilling*. Testens aldersgruppe er fra 4 til 14 år. Testen krever ingen verbal respons fra barnet, og kan derfor være hensiktsmessig i forhold til personer med ulike språkvansker (Lyster 2007). Det stilles altså ikke krav ekspressiv språkutvikling, det vil si at *testen måler grammatiske strukturer, reseptivt*.

### 7.1.1 Hva måler testen

Testen er delt opp i 20 blokker, hvor hver blokk består av fire setninger som har lik grammatisk og/ eller leksikalsk struktur. Testmaterialet består av 80 oppgaver, hvor hver oppgave representerer et ark, med fire ulike bilder. Barnets oppgave blir å peke ut korrekt bilde, ut i fra stimuli setningen som leses opp. Testleder sier en setning, f. eks: ”Sauen løper”, hvorpå barnet skal peke ut riktig bilde, av de fire bildene som presenteres. Barnet må svare riktig på alle fire oppgaver innenfor en blokk, for å oppnå skåre. Hovedformålet med utviklingen av TROG 2 var at testen spesielt skulle ta hensyn til grammatisk struktur (Bishop 1989). Og kan derfor være et godt mål hvis man har antakelser om at barnet har *underliggende mangler rent grammatisk/morfologisk*. Barns evne til å forstå språk, kan hemmes av ulik problematikk, begrenset ordkunnskap kan være en årsak, svakt verbalt minne en annen, eller grammatiske begrensinger. Derfor er TROG 2 et godt supplement dersom man har antakelser om at barnet har vansker med grammatiske strukturer. Det påpekes i manualen at et svakt resultat på TROG 2, kan forklares med at barnet har et dårlig vokabular. Dette gjelder særlig når testen brukes på barn med et annet morsmål enn norsk. Manualen viser til er et svakt punkt ved testen, fordi den viser til at vokabular har en innvirkning på resultatet til barnet. Dermed kan man si at testen i tillegg til å måle grammatisk utvikling, inntil en viss grad også gir et mål på barnets ordkunnskap.

I den engelske manualen kommer reliabilitet og validitetsaspektet noe uklart frem. Det vises blant annet ikke til at det er foretatt noen form for statistiske metoder i forhold til disse begrepene. Dette gjelder også den norske manualen.

### **7.1.2 Testens kulturelle kontekst og opprinnelse**

Testen er utviklet i Storbritannia. Det er ikke forsøkt å gjøre standardiseringen representativ for hele populasjonen i Storbritannia (Bishop 1989). I den norske manualen som er under utvikling, vises det til at testen er tilpasset norske forhold (Lyster 2007).

### **7.1.3 Testens normeringsutvalg**

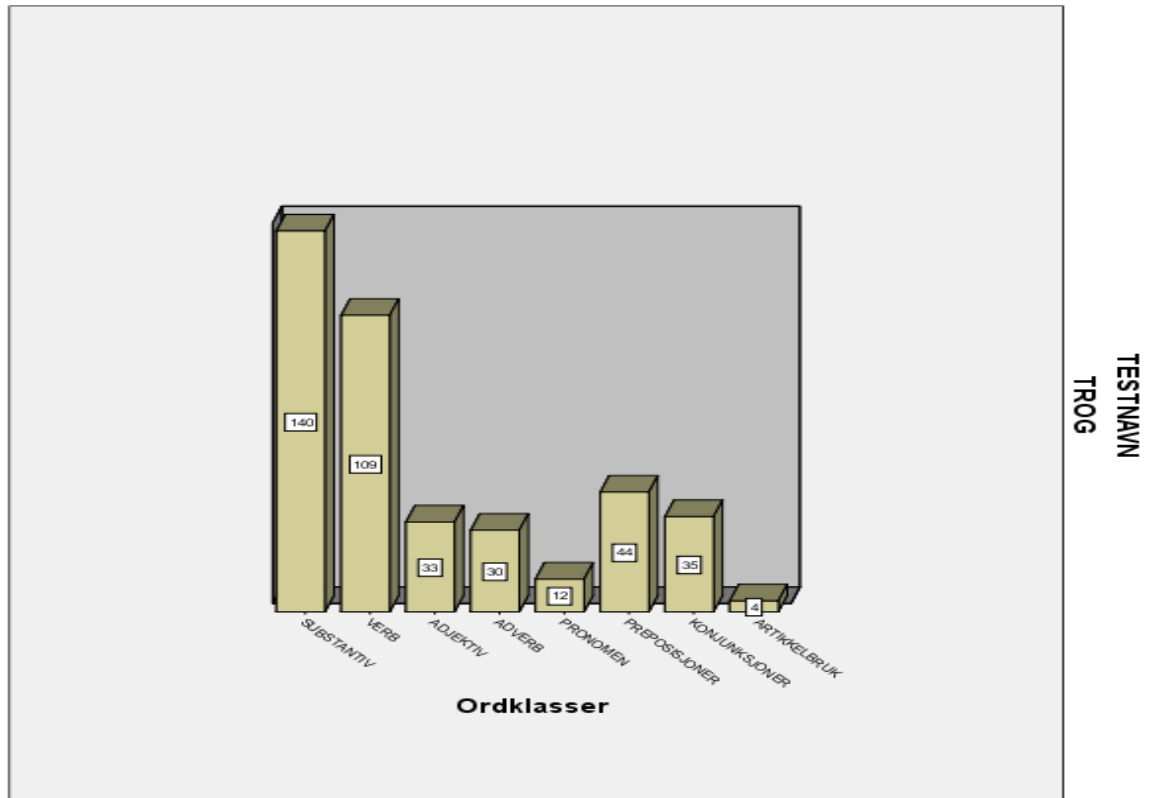
Standardiseringen i Storbritannia er foretatt på over 2000 engelsk språklige barn i alderen fire til tolv år. En rekke kriterier ble satt i forhold til hvilke barn som deltok i standardiseringen. Blant annet ble barn med foreldre som ikke hadde engelsk som morsmål ekskludert (Bishop 1989). Originalutviklerne har med andre ord ikke tatt minoritetsspråklige barn med i betraktningen under utviklingen av TROG 2.

Den norske datainnsamlingen rundt TROG 2, er foretatt i Nord-Norge, Trøndelag, Sør-Norge, samt Vestlandet og Østlandet. Det er en rekke barn som er utelatt i standardiseringen, blant annet barn som vokser opp i en familie som ikke har norsk som daglig tale. Følgelig er minoritetsspråklige elever ikke tatt med i standardiseringen. Ut i fra manualen er det noe uklart om hele landet er representert. Kan man da gå ut i fra at den norske versjonen av TROG er representativ i forhold til hele den norske barnegruppa? Blant annet sier manualen ingen ting om hvorvidt samiske barn er tatt med i standardiseringen. Videre er landsoppdelingen noe uklar, først vises det til landsdeler som Nord-Norge, Vestlandet og Østlandet, samtidig som fylket Trøndelag er representert. Dette er en uklar fremstilling, som kan gjøre testens normeringsutvalg lite representativt, og følgelig vil også barnets resultater bli svekket i forhold til testens validitet. Minoritetsspråklige er som nevnt ikke tatt med i standardiseringen, og man bør derfor som testleder vise varsomhet ved bruk av TROG 2 på denne elevgruppen, særlig med tanke på tolkningen av barnets resultater.

### **7.1.4 Ordvalg i testen**

Figur 3 nedenfor viser en oversikt av ulike ordklasser i TROG 2.

FIGUR 3: Ordklasser i TROG 2



Som man kan se av fremstillingen ovenfor inneholder TROG 2 flest substantiv, påfulgt av verb, preposisjoner, konjunksjoner, adverb, pronomen og artikler. Ordene er nokså hverdagslige, og lite variert. De samme ordene gjentas i flere av oppgavene. Dette er et bevisst valg fra testutvikleres side, på den måten sikrer de at testresultatet ikke påvirkes av ordforståelsesproblemer (Bishop1989; Lyster 2007). TROG 2 har fokusert på å ha med ord som barnet forstår. Testen undersøker i innledningsoppgavene om barnet forstår ordene, som senere brukes i testen.

## 8. The British Picture Vocabulary Scale - BVPS II

BPVS II (The British Picture Vocabulary Scale 2.ed) er utviklet av L. M Dunn, L. M Dunn, C. Whetton & J. Burley, og måler hovedsakelig *elevens reseptive ordforråd (hearing)* vokabular for standard engelsk (Dunn, Dunn mfl 1997). Begrepet reseptivt vokabular brukes for å beskrive hvordan et individ forstår ord enten ved å lese eller å høre. Ekspressivt språk viser til hvordan individet kan produsere egne setninger, enten skriftlig eller muntlig. BPVS II måler reseptivt ordforråd, som da vil tilsi at det måler hva barnet forstår, men ikke nødvendigvis kan klare å uttale eller bruke.

BPSV II er beregnet for barn fra 2 ½ til 18 år. Testen er delt opp i fjorten blokker, hver blokk inneholder tolv oppgaveitems, tilsvarende 168 oppgaver. Testen administreres ved at testleder sier et stimuliord, for eksempel: "baby", hvorpå barnet peker ut riktig bilde av de fire som presenteres. Testen har ulikt startpunkt i forhold til barnets alder. Barnet kan ikke ha mer enn en feil i startsettet, dersom dette forekommer må man reversere til settet før. Stoppregel er åtte eller flere feil innen for et sett (ibid).

BPVS har likheter med Peabody Picture Vocabulary test. PVVT måler språkevne, verbal forståelse, vokabular ferdigheter, reseptivt språk, gjenkjennelses vokabular, verbal intelligens, ordforståelse, bruk av vokabular (Sattler 2001). BPVS er nylig omarbeidet fra engelsk til norsk av Institutt for spesialpedagogikk ved Universitet i Oslo.

### 8.1.1 Hva måler testen?

BPVS II måler som nevnt hovedsakelig reseptivt ordforråd. Det betyr at testen måler de ordene som barnet forstår muntlig, det stilles altså ingen krav til å kunne uttale ordet. Den engelske manualen viser at testen i tillegg kan brukes som screening av skolefaglig tilegnelse (Dunn, Dunn mfl 1997). Dette gjelder kun dersom barnet kommer fra en engelskspråklig familie (ibid). Det betyr at det ikke er aktuelt å bruke testen i det henseende når det gjelder minoritetsspråklige elever.

Reliabiliteten i BPVS II er relativ høy, og man kan derfor anta at dersom testinstruksjonene følges, vil testen måle det samme gjentatte ganger. Validitetsaspektet viser at BPVS II har en god innholdsvaliditet, det vil si at BPVS er et godt mål på reseptivt ordforråd. Testen

---

korrelerer med en rekke andre tester som sies å måle det samme, og en kan derfor si at testen har en god begrepsvaliditet (Dunn & Dunn mfl 1997).

### **8.1.2 Testens kulturelle kontekst og opprinnelse**

Testen er utviklet i Storbritannia. Dermed kan det ses som naturlig at testens kulturelle kontekst er britisk og vestlig. Testen kan brukes til å avdekke hvorvidt minoritetsspråklige elever har tilstrekkelig kunnskap i engelsk for å kunne følge undervisning på engelsk språklige skoler og universiteter. Den norske utgaven av BPVS II er en direkte oversettelse. Det betyr at man har tatt utgangspunkt i de samme ordene på norsk. Hvor kjent ordene i testen er for barnet har betydning hvilken skåre barnet kommer ut med. Frekvensen av ordene som brukes i testen er et vesentlig moment. Høyfrekvente ord på engelsk trenger ikke nødvendigvis være høyfrekvente på norsk. Dette kan få betydning for testens vanskegrad, og kan innebære at den norske utgaven har vesensforskjeller som gjør testen vanskeligere på norsk. Dette kan ha ført til at en rekke psykometriske egenskaper ved testen kan ha falt bort under oversettelsesprosessen. Ved en oversettelse er det flere problemer som kan oppstå, blant annet sammenlignbarheten av normer og ekvivalensen av skårene. Det vil si at skårene er likeverdige, at oversettelse av testen gir samme resultat som original utgaven. Ofte er ikke en direkte oversettelse nok, noe tilpassning og revisjon er som regel nødvendig. Det kan også oppstå kulturelle forskjeller mellom itemene. Derfor skal man ikke gå ut i fra det finnes et sammenligningsgrunnlag mellom original testen og oversettelsen av den (Anastasi & Urbina 1997).

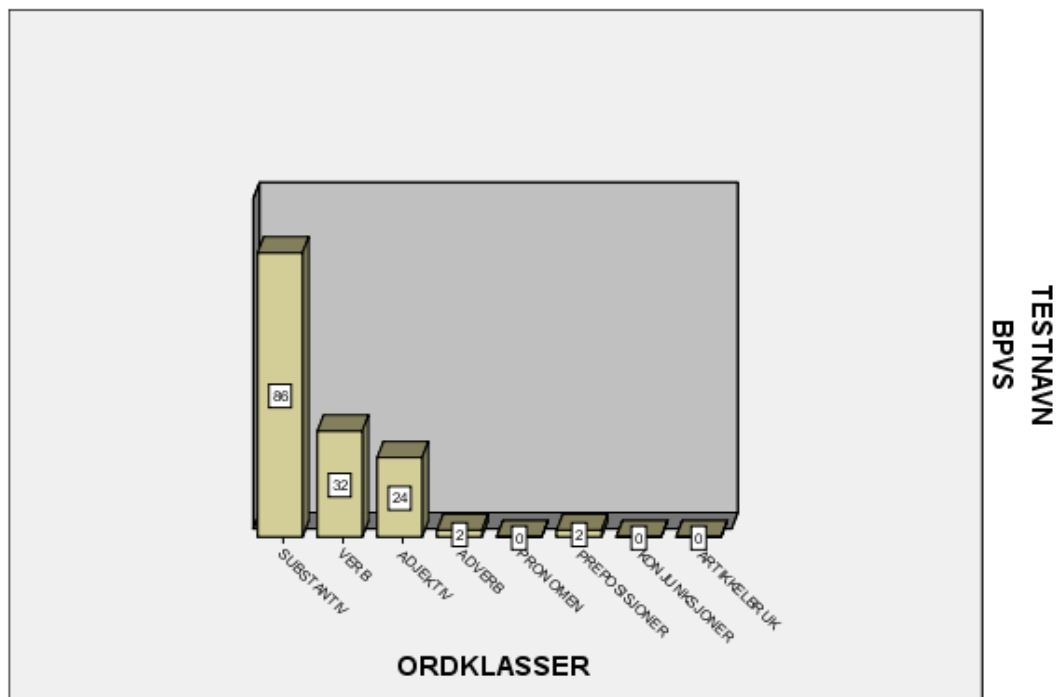
### **8.1.3 Testens normeringsutvalg**

Standardiseringen av BPVS II forgikk i England og Wales i 1995. I alt ble 2571 elever testet fra 152 ulike skoler. Standardiseringsutvalget er i følge manualen svært representativt i forhold til den nasjonale befolkningen (Dunn, Dunn mfl 1997). En rekke statistiske metoder ble brukt under standardiseringen, blant annet klassisk item analyse, item diskriminering og testens indre reliabilitet (ibid).

### **8.1.4 Ordvalg i testen**

Figur 4 nedenfor viser en oversikt av ulike ordklasser i BPVS II.

FIGUR 4: Ordklasser i BPVS II



Som man kan se av oversikten over inneholder BPVS II; flest substantiv som de to andre testene. Deretter følger verb, adjektiv og adverb. BPVS II er den testen som inneholder færrest ordtyper.

Den engelske manualen påpeker at ordvalgene i BPVS II er valgt på grunnlag av variasjon, samt en balanse av ulike ordtyper. Ordtypene er klassifisert i 29 grupper, eksempler på dette er blant annet handlinger, adjektiver, deler, mennesker, mat og så videre. Stimuli ordene har i stor grad en forbindelse til hverdagspråk, men for å skape øvre grense for testen, har det blitt tatt inn utvalgte ord som er noe mer avansert og mindre brukt til dagligdags (Dunn, Dunn mfl 1997).

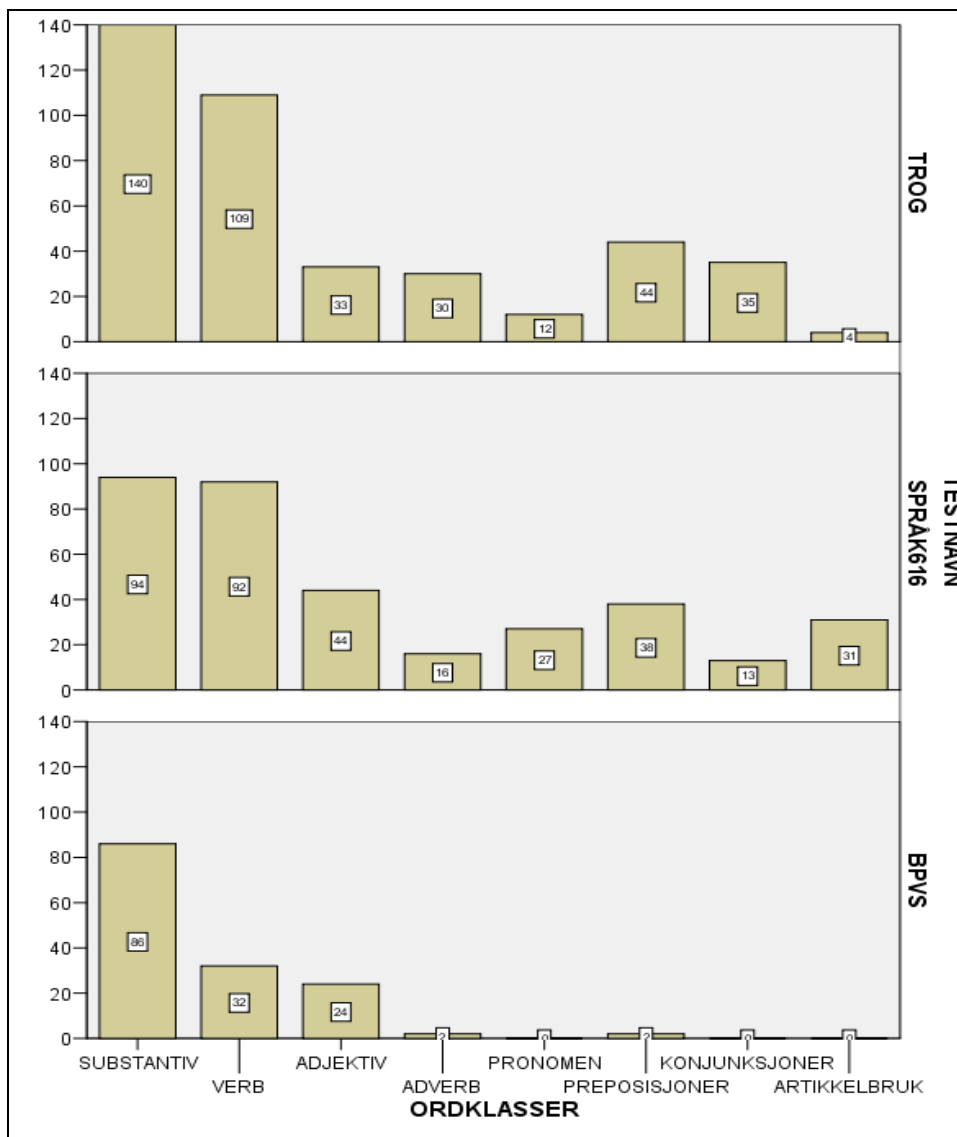


## 9. Sammenligning

Språk 6 – 16 er betraktelig mer utfyllende for kartlegging av språk enn BPVS II og TROG 2. Testen måler blant annet flere aspekter ved språket, og har som sagt antakelser om at det finnes underliggende separate kognitive ferdigheter som måles av testen (Ottem & Frost 2005). For det første inneholder testen flere deltester enn TROG 2 og BPVS II. Dette er også testen som har flest ordtyper. Antakelsen om at det er en sammenheng mellom den fonologiske løkken og arbeidsminnet som gjør at enkelte barn får problemer med språktilegnelsen viser at Språk 6 – 16 har flere psykometriske egenskaper enn TROG 2 og BPVS II.

### *Prosentandel av antall ord, fordelt på ulike ordklasser*

|                      | <b>TROG 2</b> | <b>BPVS II</b> | <b>Språk 6 – 16</b> |
|----------------------|---------------|----------------|---------------------|
| <b>Substantiv</b>    | 34,39 %       | 59,72 %        | 24,86 %             |
| <b>Verb</b>          | 26,78 %       | 22,22 %        | 24,33 %             |
| <b>Adjektiv</b>      | 8,10 %        | 16,66 %        | 11,64 %             |
| <b>Adverb</b>        | 7,37 %        | 1,38 %         | 4,23 %              |
| <b>Pronomen</b>      | 2,94 %        | 0              | 6,87 %              |
| <b>Preposisjoner</b> | 10,81 %       | 0              | 10,05 %             |
| <b>Konjunksjoner</b> | 8,59 %        | 0              | 3,43 %              |
| <b>Artikkelbruk</b>  | 0,98 %        | 0              | 8,2 %               |
| <b>Spørreord</b>     | 0             | 0              | 6,34 %              |

**FIGUR 5: Sammenligning av ulike ordklassene i de analyserte testene.**

Som man kan se av oppsummeringen ovenfor inneholder alle testene flest substantiv. BPVS II inneholder færrest ordtyper, og dette kan forklares gjennom at testen er oppbygd slik at barnet kun får presentert et ord, for hver oppgave som gis til barnet. Dette begrenser antall ordtyper, og testutviklerne har for det meste tatt for seg substantiv og verb for å beskrive bildene som vises til barnet.

Språk 6- 16 er mer variert med tanke på ordvalg i testen. Det er flere årsaker til dette. For det første inneholder Språk 6- 16 flere varierte oppgaver enn TROG 2 og BPVS II. For det andre måler Språk 6 – 16 som nevnt, ulike aspekt ved språket enn det de to andre testene har til hensikt å måle. Barnet blir blant annet bedt om å gjenta setninger med økende lengde, dette betyr økt antall ord, og større variasjon i ordtypene.

---

TROG 2 inneholder også et variert ordvalg. Det påpekes i manualen, at de første oppgavene i testen er til for å sikre at barnet forstår alle ordene som brukes senere i testen.

### 9.1.2 Bildene i TROG 2 og BPVS II

Både TROG 2 og BPVS II består av bilder. Er det plausibelt å gå ut i fra at bildefremstillingen og valg av bilder har betydning for barnets skåre, og i hvilken grad er det påvirkbart? Representerer bildene en spesifikk kultur eller er bildene fri for kulturelle bias? Er det noen av bildene som representerer minoritetsspråklige barns kultur?

TROG 2 består av 80 bilder. Manualen viser til at bildene som er tatt i bruk i TROG 2 er klare og tydelig bilder, som er utviklet for at barnet ikke skal kunne la seg påvirke av ikke lingvistiske bias under selve testingen. Etter å ha gjennomgått bildene i TROG 2 vil jeg påpeke en rekke ting i forhold til bildene som er tatt i bruk. For det første kan det se ut til at bildene representerer vestlig kultur. De fleste menneskene som finnes på bildene har hvit hud. Dyrene kan ses på som vanlige dyr som finnes på en bondegård i vesten, som for eksempel: ku, sau, and, hest, hund og katt. Av de 80 bildene som forekommer i testen, inneholder 10 av disse mennesker med en annen hudfarge enn hvit. Når det gjelder antall jenter og gutter på bildene er dette jevnt fordelt, med 100 jenter og 104 gutter. Det er ikke foretatt noen endringer på bildene i TROG 2 og en må derfor gå ut i fra at testutviklerne går ut i fra at det er samme kulturelle forhold i Norge som i Storbritannia. Dette kan ses på som en svakhet ved testen.

BPVS II inneholder 168 oppgaver, med fire bilder i hver oppgave. Bildene i BPVS II er nokså nøytrale i sort/hvitt. Til tross for dette kommer det klart frem av bildene hvem som har minoritetsspråklig bakgrunn. Av de 672 bildene er det 96 som har innhold som kan representere en form for minoritetsspråklig bakgrunn. Antall jenter og gutter i BPVS II, er noe ujevnt. Da det forekommer 44 flere bilder av gutter. Dette er en noe ubalansert fremstilling, og kan være en svakhet ved testen. Bildene i BPVS II kan sies å være mer variert enn bildene i TROG 2. Bildene inneholder flere ulike mennesker, dyr, hus, situasjoner og steder enn det som vises på bildene i TROG 2.

Legger man bildetyperne til grunn kan man si at BPVS II i større grad enn TROG 2 er representativ i forhold til kartlegging av norskspråklige ferdigheter hos minoritetsspråklige elever.

## 10. Refleksjoner over funn fra analysen

I forhold til oppgavens problemstilling viser det seg at TROG 2, BPVS II og Språk 6 – 16 kan være lite hensiktsmessige som kartleggingsverktøy for norskspråklig utvikling hos språklige minoriteter. Manualene til TROG 2 og Språk 6 – 16 påpeker at testene ikke bør brukes på denne elevgruppen. Dette sett i sammenheng med at det minoritetsspråklige ikke er tatt med i normeringen, ser en antydning til at dette kanskje ikke er den mest hensiktsmessig kartleggingen av minoritetsspråklige elevers norskspråklige utvikling.

Original utgaven av BPVS II har med en normering med tanke på andrespråkselever, men dette er kun en vurdering av hvorvidt eleven har gode nok ferdigheter til å følge engelsk språklig undervisning (Dunn & Dunn 1997). Den norske videreutviklingen har ikke tatt med minoritetsspråklige i standardiseringen. Dette gjelder samtlige av testene som er analysert.

Dersom TROG 2, BPVS II og Språk 6 – 16 skal brukes på minoritetsspråklige elever vil dette kunne svekke testenes validitet og reliabilitet. Om testene skulle tas i bruk på denne elevgruppen, er det nødvendig at testleder er klar over den manglende standardiseringen, og at det tas hensyn til dette under tolkningen av barnets resultat. Under kartlegging av et barn, er det nødvendig å samle inn informasjon fra flere informanter, samtidig er flere tester påkrevet for å kunne komme frem til adekvat kartlegging av barnet. Det som vanskeliggjør saken i minoritetsspråklige barns tilfeller, er at det ikke finnes noen kartleggingsverktøy for norskspråklig utvikling.

Det kan ses på som bekymringsfullt at Bredtvet og Torshov kompetanse senter utvikler Språk 6 – 16 til bruk på minoritetsspråklige barn. Dette er en test som i utgangspunktet er tiltenkt norskspråklige barn. Hvordan kan dette være et godt utgangspunkt for kartlegging av minoritetsspråklige barns norskspråklige utvikling? Dette gjelder for så vidt også TROG 2 og BPVS II, den norske oversettelsen er ikke tiltenkt språklige minoriteter. Hva bør fokusområdene være for at TROG 2, BPVS II og Språk 6 – 16 skal bli gode kartleggingsverktøy for norskspråklig utvikling hos språklige minoriteter? Første prioritet er at testene standardiseres i forhold til denne gruppen. Dersom det lages et normeringsgrunnlag som er tilpasset denne gruppen, vil testene kunne komme frem til mer reliable og valide resultater for hvert enkelt barn. Språk og kulturelle betingelser bør bli tatt med i betraktning, i en videreutvikling av disse testene til bruk på minoritetsspråklige elever. Valgene som foretas bør inkludere en nøye gjennomtenkning av ord og kulturelle innsalg, og

---

dermed være minst mulig påvirket av språklige og kulturelle forskjeller. Dersom man ønsker å undersøke hvorvidt barn med minoritetsspråklig bakgrunn har språkvansker, eller kun manglende ferdigheter i norsk, kan det som nevnt tidligere være nødvendig å utvikle parallelle tester på barnets morsmål.

Analysen viser at samtlige av testene inneholder flest substantiv. Språk 6 – 16 inneholder flest ulike ordklasser. Dette kommer av denne testen undersøker flere aspekter ved språket, og er en mer utfyllende test enn BPVS II og TROG 2. BPVS II og TROG 2 er direkte oversettelser. Dette kan som nevnt ses på som negativt i den retning av at visse psykometriske egenskaper ved testen kan ha falt bort under oversettelsen. En oversettelse av en test bør medføre visse endringer, en ren oversettelse er som regel ikke dekkende. Fordi det kan finnes både kulturelle og språklige betingelser som kan være forskjellig innenfor de to ulike språkene (Anastasi & Urbina 1997). Dette kan få betydning for testens psykometriske egenskaper, som for eksempel testens reliabilitet og validitet. Samtidig kan vanskegraden på ordene ha blitt vesentlig høyere etter en oversettelse. Dette kommer av at engelsk er et mer innholdsrikt språk enn norsk, som betyr at ord som er høyfrekvente på engelsk, ikke nødvendigvis er like høyfrekvente på norsk.

Bildene i TROG 2 og BPVS II, er de samme bildene som er brukt i original testene. Hvorvidt dette er problematisk er vanskelig å avgjøre, men norsk og engelsk har noe av den samme vestlige kulturen, så bildene trenger ikke være påvirket av kulturelle betingelser som utgjør vansker for norske barn. Men minoritetsspråklige elever på sin side, kan ha ulik kulturell bakgrunn enn fra det som representert gjennom bildene, som kan påvirke deres resultater på testen.

En del av ordene i Språk 6 – 16, kan ses på noe utdaterte, ordet: "Fure" i deltest 4 *Ordkunnskap* er et eksempel. Dette er et ord som trolig få barn i aldersspannet 6 – 16 har hørt om. Derfor er det angivelig å tro at dette også gjelder minoritetsspråklige elever. Enkelte av setningen som skal gjentas i deltest 1 *Setningsminne*, er setninger som inneholder kulturelle bias. Som nevnt er setningen: "Mor og far reiste ut på tur i en liten kano" preget av norsk kultur. Dette kan også sies å være et norsk middelklassefenomen. Å gå på tur, er et nokså særegent norsk middelklasse fenomen. Mange minoritetsspråklige har lite erfaring med å dra ut på kanotur. Derfor kan dette tolkes dit hen at Språk 6 – 16 er lite hensiktsmessig å bruke på språklige minoriteter, fordi det både er kulturelle og språklige betingelser, som ligger til grunn for at resultatene blir lite valide.

Det har ikke blitt foretatt noen analyse av WISC III, men denne testen er tatt med som et eksempel på en test som inneholder både språklige og kulturelle betingelser, som betyr at man skal vise varsomhet ved tolkning av minoritetsspråklige resultater. Det er særlig de språklige oppgavene hvor det har vist seg at enkelte minoritetsspråklige barn får betraktelig dårligere resultater, enn majoritetsspråklige barn. Dette peker i retning at andre språktester også kan ha de samme kulturelle og språklige betingelsene som betyr at man skal vise varsomhet ved tolkning av resultater hos denne elevgruppen.

Språklige minoriteter vil i mange tilfeller ha andre og ulike læringserfaringer enn majoritetsspråklige elever, dette utgjør store forskjeller mellom disse to gruppene. Dette kan bety store variasjoner i resultatene på ulike tester. TROG 2, BPVS II og Språk 6 – 16 er utviklet for barn med læringserfaringer fra norsk kultur og språk, og vil derfor ikke være hensiktsmessig å bruke på minoritetsspråklige. Oppsummert skal det sies at de analyserte testene inntil en viss grad reflekterer norsk språk og kultur. Dette kan få betydning for minoritetsspråklige elevers resultater, som oftest i negativ retning. Om testene gir et bilde av minoritetsspråklige elevers norskspråklige ferdigheter er heller usikkert, fordi det er så mange faktorer som spiller inn med tanke på resultatet. Bildene, de språklige formuleringene i testene, språklige ferdigheter, kulturelle og erfaringsmessige faktorer hos elevene utgjør forskjeller som kan prege resultatene til minoritetsspråklige elever på de nevnte testene.

## 11. Pedagogiske implikasjoner

Denne oppgaven har først og fremst blitt skrevet for å sette lys på at det fremdeles er mangel på ulike kartleggingsverktøy for språklige minoritetselever i Norge. Dette er noe som pekes på fra en rekke andre hold. Blant annet viser en evaluering utført av Rambøll management (2006) at det er mangel på kartleggingsverktøy i forhold til språklige minoriteter i Norge. Det som har vært utgangspunktet for oppgaven er å få frem at det er et behov på slike tester. De testene som er under utvikling i dette henseende, kan vise seg i ettertid og ikke være like godt verktøy som tenkt. Nettopp fordi de i utgangspunktet ikke har vært tiltenkt denne elevgruppen eller norske elever for den sak skyld. Konsekvensene av dette kan føre til at testene i utgangspunktet er lite hensiktsmessig for den type kartlegging, og at en oversettelsesprosess gjør testen vesentlig mye dårligere. De testene som er analysert i denne oppgaven, er alle tester som viser til kartlegging av ulike delferdigheter ved språkutvikling og/eller ulike aspekter språkutvikling.

Jeg ønsker at analysen, perspektivene og drøftingen i denne oppgaven kan kunne bidra til utvikling av slike tester. Som nevnt tidligere er det vanskelig å utvikle en kulturfri test, så hvordan kan man teste minoritetsspråklige elever, og komme frem til valide og reliable resultater. For det første er det viktig å kunne ta med barnets morsmål i betraktning i tolkningen av resultatet. At morsmålet ikke er det samme som majoritetsspråket, og kan være en årsak til et eventuelt svakt resultat.

Testene som skal brukes for å identifisere elever med språkvansker, enten av spesifikk art eller manglende norsk språklige ferdigheter, bør muligens utvikles slik at pedagoger som kjenner eleven kan utføre testingen. Slik vil man unngå at eleven utsettes for en uvant situasjon som igjen kan påvirke resultatet. En som jobber med eleven daglig, vil også ha egne antakelser om hva barnets vansker er.

Testene som utvikles bør ha egne normer for minoritetsspråklig elever. Dette er særlig viktig for de elevene som er svært svake. Minoritetsspråklige elever som mestrer det norske språk, kan eventuelt sammenlignes på lik linje som majoritetsspråklige elever. Den svakeste gruppen av minoritetsspråklige elever burde sammenlignes med et representativt utvalg for denne elevgruppen. Det viser seg at samtlige tester som er i bruk i dag mangler normeringsgrunnlag for språklige minoriteter.

Vokabular kan sies å være den viktigste faktoren med hensyn til språkutvikling og skolesuksess (Vermer 2001). Andrespråk innlæring har først og fremst blitt målt gjennom bredde vokabular, (hvor mange ord kan de) enn dybde (hvor godt kjenner de ordene). Det å bare kunne gjenkjenne ordene er omdiskutert, fordi det vanskelig lar seg avgjøre i hvilken grad ordene gjenkjennes (ibid). Testene som er diskutert i denne oppgaven er tester som i stor grad kun undersøker breddevokabularet til barnet. Spesielt BPVS II gir et mål på bredde vokabular, som kan ses på som en svakhet fordi det vanskelig lar seg avgjøre i hvilken grad barnet gjenkjenner ordet. Som testleder må man gå ut i fra at barnet kun gjenkjenner ordet inntil en viss grad. Hvor godt ordet gjenkjennes av barnet, kan være vanskelig å teste gjennom de testene som er beskrevet i denne oppgaven. Kanskje det er på tide å utvikle en test som undersøker barnets dybdevokabular?

Dagens praksis vedrørende andrespråksinnlæring og opplæringen av språklige minoriteter er mangelfull, og kan ofte være lite hensiktsmessig rent pedagogisk for hvert enkelt barn det måtte gjelde. Øzerk (2007) har gjennom sine undersøkelser avdekket en rekke mangler ved dagens ordning. Det gjelder først og fremst kartleggingsprosedyrene som ikke er standardisert, og at skolene mangler retningslinjer når det gjelder testing av denne elevgruppen. Samtidig som avgjørelsen om hvorvidt eleven har krav på særskilt norskopplæring hovedsakelig baseres på en skjønnsmessig vurdering foretatt av lærere rundt den aktuelle eleven (Øzerk 2007). En videreutvikling av denne praksisen ses på som uheldig. Min oppgave kan bidra til å belyse og formulere nye problemstillinger rundt dette området. Slik at man kan komme frem til en mer formålstjenelig ordning rundt kartleggingsprosedyrer og andrespråksopplæring.

Analysen viser at testene som er under videreutvikling av Bredtvot og Torshov kompetanse senter, for det første ikke er tiltenkt språklige minoriteter fra original utviklerne sin side. Å bygge videre på disse testene kan være lite hensiktsmessig, da det kan være kulturelle og språklige betingelser, som utgjør signifikante forskjeller, hvilket betyr at resultatene til minoritetsspråklige blir lite valide.

WISC III, er tatt med i oppgaven for å beskrive en test som er oversatt og videreutviklet på svensk, for så å bli tatt i bruk her i Norge. WISC III er som nevnt en test som har blitt usatt for en del kritikk i forhold til bruk på minoritetsspråklige elever i Norge. WISC III har kulturelle og språklige innslag som kan påvirke resultatene til språklige minoriteter i negativ retning. Det samme gjelder BPVS II og TROG 2, dette er tester som er utviklet i



---

Storbritannia og oversatt og tilpasset norske forhold. Det er viktig å stille spørsmålstegn ved i hvilken grad disse testene er tilpasset norske forhold. Det ser ut til at tilpasset til norske forhold betyr en direkte oversettelse fra engelsk til norsk, samt en norsk normering. Hvorvidt en slik videreutvikling gjensker de samme psykometriske egenskapene som testens originalutvikler hadde, er vanskelig å svare på, men det er viktig å være klar over at endringer av testens psykometriske egenskaper kan forekomme. Som sagt kan det forekomme endringer av sammenligningsgrunnlaget. Videre bør det foretas en grundig gjennomgang av oversettelsen, denne prosessen krever som regel tilpasning og revisjon i forhold til både språklige og kulturelle forskjeller som kan oppstå mellom ulike språk (Anastasi & Urbina 1997). Reliabilitet, validitet og normer i oversettelsen bør utvikles i forhold til den populasjonen testen skal brukes i (ibid). Dette viser igjen hvor viktig det er å utvikle normer i forhold til minoritetsspråklige elever i Norge.

Samme kritikk som WISC III blir utsatt, gjelder også Språk 6 – 16, TROG 2 og BPVS II, siden dette er rene språktester, dette kan by på vanskeligheter for språklige minoriteter. Derfor bør man som testleder vise varsomhet i forhold til tolkning av språklige minoriteters resultater på disse testene. Slik kan man unngå feildiagnostisering, feilplassering av elever i forhold til spesialundervisning og at minoritetsspråklige elever blir ansett for å ha svake evner.

De svake sidene til testene som er analysert kan oppsummeres følgende, testene har ikke en standardisering i forhold til språklige minoriteter, dette svekker testenes reliabilitet og validitet. Testene inneholder også språklige, kulturelle og erfaringsmessige bias som kan slå uheldig ut med tanke på minoritetsspråklige elevers resultater. Både TROG 2 og BPVS II er rene oversettelser fra engelsk, dette kan ha betydning for testens reliabilitet og validitet. En slik oversettelse vil kunne ha betydning i forhold til hvilke ord som tas med, dette sett i sammenheng med begrepene hverdagsspråk og akademisk språk. Det er vanskelig å avgjøre i hvilken grad testene måler begge språkaspektene, eller kun et av de. I forhold til skolegang ville det i stor grad være mest hensiktsmessig om testene målte akademisk språk, som i større grad vil kunne predikere og vise til videre skolegang hos eleven.

De testene som er analysert i denne oppgaven, er alle tester som viser til kartlegging av ulike delferdigheter ved språkutvikling og/eller ulike aspekter ved språkutvikling. Testene er enkle og raske å gjennomføre. En lærer/pedagog vil kunne gjennomføre testene på en adekvat måte. Dette er positivt, fordi det viser seg som nevnt tidligere at skolene mangler

kartleggings rutiner og adekvate kartleggingsverktøy. Dersom TROG 2, BPVS II og Språk 6 – 16 får normeringsutvalg som inkluderer språklige minoriteter, vil det bli mer naturlig å bruke disse testene på denne elevgruppen.

Reflekterer testene et mål av elevens nåværende nivå i norsk? Dette er et viktig spørsmål som en bør stille seg ved bruk av testene. Samlet sett kan testene gi et bilde av elevens språkutvikling, og nåværende nivå i norsk. De ulike testene måler alle forskjellige aspekter ved språket. TROG 2 måler grammatisk utvikling, BPVS II gir et bilde av elevens reseptive ordforråd, men Språk 6 – 16 undersøker det språklige korttidsminnet samt utviklingen av språklige begreper (Ottem & Frost 2005).

### **11.1.1 Hva er alternativet?**

Spørsmålet er svært aktuelt, for per dags dato finnes det ingen alternativer til kartlegging av minoritetsspråklige barns norsk språklige ferdigheter eller evner. Derfor brukes det tester som egentlig ikke er tiltenkt denne gruppen, i mangel og påvente av noe annet. Bredtvet og Torshov kompetanse senter driver som sagt med videreutvikling av TROG 2, BPVS II og Språk 6- 16. Prosjektet er å oversette disse testene til diverse morsmål, slik at morsmålet kan kartlegges. Det blir foretatt en normering i forhold til minoritetsspråklige. Dette er et skritt i riktig retning. Det er gjennom egenutviklede tester en kan kartlegge på en adekvat måte. Men hvor egenutviklet er en videreutvikling av TROG 2 og BPVS II, som opprinnelig ikke er tiltenkt språklige minoriteter?

Det som er problematisk i forhold til minoritetsspråklige barns språkferdigheter er at denne gruppen er langt fra homogen, det betyr at det finnes et flertall av språklige og kulturelle ulikheter blant språklige minoriteter som gjøre det vanskelig å lage en test, som kan dekke hele gruppen av minoritetsspråklige barn.

Det ses som positivt å sette i gang tiltak på et tidlig tidspunkt. Dette er svært avgjørende for barnets videreutvikling av norsk språklige ferdigheter. Desto tidligere det iverksettes tiltak med tanke på språkutviklingen, jo bedre resultat for eleven. Utvikling av morsmål ses også som en positiv faktor med tanke på andrespråksinnlæring, det er derfor lite hensiktsmessig at en slik utvikling ikke legges vekt på i skolesammenheng. I følge Cummins vil tospråklighet fremme lingvistisk, kognitiv og akademisk utvikling når begge språk får mulighet til utvikling (Cummins 2000). For å utvikle et tilbud til barnet, selv innenfor rammen av tilpasset opplæring, er det nødvendig å avdekke konkret hva barnets vansker innebærer,

samtidig som man utvikler et opplæringstilbud som på lengre sikt kan hjelpe barnet til å utvikle norskspråklige ferdigheter. Det påpekes igjen at forskning viser at barnet langt på vei kan hjelpes til norskspråklig utvikling, dersom morsmålet utvikles på en adekvat måte (Øzerk 2003, Cummins 1979). Det legges ikke særlig vekt på morsmåls utvikling i dagens læreplan, noe som kan anses som uheldig for enkelte minoritetsspråklige elever.

Alternativt vil det være svært fordelaktig dersom det utvikles parallelle tester på barnets morsmål, slik kan man som nevnt kartlegge barnets morsmål og andrespråk på en adekvat måte. Slik vil man kunne avdekke hvorvidt barnet har språkvansker eller kun manglende ferdigheter i norsk. Dersom barnet har språkvansker vil språkvanskene være synlig både på barnets morsmål og i norsk.

## 12. Avsluttende kommentarer

Minoritetsspråklige elever øker i antall, særlig i Osloskolen. Derfor vil det bli nødvendig å utvikle gode kartleggingsrutiner som hver enkelt skole kan bruke. Det er vesentlig at kartleggingen er lik over i hele landet slik at det ikke forekommer noen differensiering mellom elevene, og da særlig i forhold til hvilke elever som får særskilt norsk opplæring og spesial undervisning. Etterspørselen etter passende kartleggingsverktøy er stor, fra skoler og lærere, men også fra Kunnskapsdepartementet.

Målet med analysen var å undersøke hvorvidt Språk 6 – 16, TROG 2 og BPVS II er gode kartleggingsverktøy med tanke på norskspråklig utvikling hos språklige minoriteter. Har jeg fått svar på det jeg spurte etter? På en måte har jeg fått svar på det jeg ønsket å undersøke, testene viser seg å ikke være det mest passende kartleggingsverktøyet. Men dersom de brukes med varsomhet, samtidig som testleder er klar over at det ikke finnes et normeringsgrunnlag for språklige minoriteter, kan disse testene være et alternativ. Kanskje også det eneste alternativet. Gjennom analysen har jeg sett på en rekke punkter ved de enkelte testene, både sterke og svake sider ved testene. Det mest hensiktsmessige vil være å utvikle parallelle tester på barnets morsmål. Dette vil kunne føre til en adekvat kartlegging av barnets morsmål, samt norsk. Dette vil kunne bidra til å undersøke hvorvidt barnet har mangelfulle ferdigheter i norsk eller språkvansker.

Testene som har vært under analysering i denne oppgaven, er ulike tester som alle har til hensikt å måle en side ved språket. TROG 2 ser hovedsakelig på grammatiske ferdigheter hos barnet, BPVS II måler barnets reseptive ordforråd, mens Språk 6 – 16 undersøker barnets språklige kortidsminne samt utvikling av språklige begreper. Men vil disse tre testene sammen kunne reflektere et mål på elevens nåværende nivå i norsk? Dersom de gjør det, samtidig som det finnes normeringsgrunnlag som har tatt med språklige minoriteter, kan dette være gode kartleggingsverktøy for kartlegging av norsk språklige ferdigheter hos minoritetsspråklige elever i Norge.

I hvilken grad kan man forsvare bruken av de analyserte testene? Hvor gode verktøy er de i forhold til å avdekke norskspråklig utvikling hos minoritetsspråklige barn? Hver og en av testene har positive sider. Men det er et vesentlig moment som testene har til felles, ingen av testene er utviklet med tanke på kartlegging av minoritetsspråklige barns språklige ferdigheter, verken originalutgaven, eller de videreutviklete norske utgavene. Dette er

kritikkverdig av følgende årsaker, for det første mangler man normeringer i forhold til språklige minoriteter. Dermed blir sammenligningsgrunnlaget feilaktig. For det andre er det et flertall av kulturelle bias i testene, som utgjør forskjeller i forhold til elever med ulik kulturell bakgrunn. Det er også grunn til å stille spørsmål rundt hvor mye disse testene brukes på språklige minoriteter? Hva er alternativet dersom disse ikke skal brukes? Hvis disse testene brukes, er det nødvendig som testleder og være klar over hvilke faktorer som kan ha betydning for barnets resultat, og ta dette med i betraktning under tolkningen av barnets resultater.

Avslutningsvis vil jeg påpeke at fremtidens elever i norsk skole, vil ha behov for ulik kartlegging. Dette vil som nevnt kreve at det utvikles parallelle tester på ulike morsmål. Det er kun slik en kan foreta en adekvat kartlegging av språklige ferdigheter hos minoritetsspråklige elever.

## Kildeliste

- Alvesson, M. & K. Sköldböck (1994). Tolkning och reflektion: vetenskapsfilosofi och kvalitativ metod. Lund, Studentlitteratur.
- Anastasi, A. & S. Urbina (1997). Psychological testing. New York, Macmillan.
- Baddeley, A, Gathercole, S & Papano C (1998). The Phonological Loop as a Language Learning Device. The American Psychological Association, Psychological Review Volume 105 (1) January 1998, side 158- 73.
- Bakken, A. (2007). NOVA Rapport 10/07. Virkninger av tilpasset språkoppl ring for minoritetsspr klige elever. En kunnskapsoversikt.
- Befring, E. (2002). Forskingsmetode, etikk og statistikk. Oslo, Samlaget.
- Bishop, D. (1989). Test for the Reception of Grammar: TROG, 2<sup>nd</sup> edn. Manchester, UK, Age and Cognitive Performance Research Centre.
- Buros, Oscar K. i Sattler, J. M. (2001). Assessment of children: behavioral and clinical applications. San Diego. (s 1) J.M. Sattler, Publ.
- B  I & Helle L. (2002) Pedagogisk ordbok: praktisk oppslagsverk i pedagogikk, psykologi og sosiologi / Oslo: Universitetsforlaget.
- Christophersen, Knut-Andreas (2004): Databehandling og statistisk analyse med SPSS. Unipub forlag. 2. utgave.
- Cronbach L (1990): Essentials of Psychological Testing, Fifth Edition. HarperCollinsPublishers. New York.
- Cummins, J. (1979). "Linguistic interdependence and the educational development of bilingual children." Review of Educational Research 49. : s. 222-251
- Cummins, J. (1984). "Bilingualism and Special Education: Issues in Assessment and Pedagogy. Multilingual Matters Ltd.
- Cummins, J. (2000). Language, power and pedagogy: bilingual children in the crossfire. Clevedon, Multilingual Matters.

---

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). The British Picture Vocabulary Scale (2<sup>nd</sup> ed). London.

Green, B. F. Jr. (1981) A primer of testing. American Psychologist, 36 s 1001 - 1011)

Helstrup, T & Kaufmann G. (2000): Kognitiv Psykologi. Bergen fagbokforlaget.

Hjardemaal, F.: Vitenskapsteori. I: Kleven, T.A. (red.): Innføring i pedagogisk forskningsmetode. En hjelp til kritisk tolkning og vurdering. (28-60) [Oslo], Unipub.

Hoskyn, M. Language Processes and Reading Disabilities. I Wong, B. Y. L. (2004). Learning about learning disabilities. (s 93 - 131) Amsterdam, Elsevier.

Kleven, T. A.: Begrepsoperasjonalisering. I Lund, T. (red): Innføring i forskningsmetodologi. (141 - 184). Oslo, Unipub.

Kleven, T. A.(2002): Innføring i pedagogisk forskningsmetode: en hjelp til kritisk tolkning og vurdering. [Oslo], Unipub.

Kvernbekk, Tone (2002): Vitenskapsteoretiske perspektiver. I: Thorleif Lund (red.): Innføring i forskningsmetodologi. (19 - 78). Oslo. Unipub.

Kunnskapsdepartementet (2007). Likeverdig opplæring i praksis!: strategi for bedre læring og større deltakelse av språklige minoriteter i barnehage, skole og utdanning 2007-2009. [Oslo], Kunnskapsdepartementet.

"[http://udir.no/upload/Satsningsomraader/UDIR\\_Likeverdig\\_opplaering\\_07.pdf](http://udir.no/upload/Satsningsomraader/UDIR_Likeverdig_opplaering_07.pdf)" (Lesedato 20nov 2007)

Kunnskapsdepartementet (2007a) - St. meld nr 16 (2006 - 2007) ... og ingen sto igjen. Tidlig innsats for livslang læring [Oslo], Kunnskapsdepartementet.

"<http://www.regjeringen.no/Rpub/STM/20062007/016/PDFS/STM200620070016000DDDDPDFS.pdf>" (Lesedato 1des 2007)

Lund, T. (2002). Innføring i forskningsmetodologi. Oslo, Unipub.

Lundgaard, Hilde for Aftenposten Aften (2008). 8 Januar Uke 2: "Er i flertall på 48 av 138 skoler"

Lyster, S (2007). Test for the Reception of Grammar: TROG, 2<sup>nd</sup>. Forløpig utgave av den norske utgave av TROG 2 fra Spesialpedagogisk Institutt, Universitetet i Oslo

Nordahl, T. and T. Overland (1998). Idealer og realiteter: evaluering av spesialundervisningen i Oslo kommune. Oslo, Norsk institutt for forskning om oppvekst, velferd og aldring.

Opplæringslova (1998): Lov om grunnskolen og den vidaregåande opplæringa av 1998. Sist endret 2008-01-01. "<http://lovdata.no/all/nl-19980717-061.html>" (Lesedato 13mars 2008)

Ottem, E. (2007). "Skolepsykologi." Språk 6- 16: Screening-test Manual II, Nye normer til bruk i forskning og utredning. (13 - 23). Tidligere utgiversted: Brandbu. Tidligere navn på utgiver: Forum for psykologer i skolen

Ottem, E. and J. Frost (2005). Språk 6-16: screening test : manual. Oslo, Bredtvet kompetansesenter.

Ottem, E, Lian, A og Karlsen P, J (2007). Reasons for the growth of traditional memory span across age. European Journal of cognitive Psychology (s. 233 - 270).

Pihl, J (2005). Etnisk mangfold i skolen. Det sakkyndige blikket. Oslo. Universitetsforlaget.

Rambøll management. November 2006. Evaluering av praktiseringen av norsk som andrespråk for språklige minoriteter i grunnskolen.

[http://udir.no/upload/Rapporter/Evaluering\\_av\\_Norsk\\_som\\_2\\_sprak.pdf](http://udir.no/upload/Rapporter/Evaluering_av_Norsk_som_2_sprak.pdf) (Lesedato 24nov2007)

Ringdal, Kristen (2007): Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode. Bergen: Fagbokforlag.

Sattler, J. M. (2001). Assessment of children: behavioral and clinical applications. San Diego, J.M. Sattler, Publ.

Sergeant J. & Taylor Eric (2002): Psychological testing an Observation. I Rutter. M & Taylor E (red): Child and Adolescent Psychiatry, 4ht Ed. (s.87 - 102) Oxford: Blackwell Publishing.

Statistisk sentralbyrå SSB. (2008):

"[http://www.ssb.nwo/emner/02/barn\\_og\\_unge/2007/bef/](http://www.ssb.nwo/emner/02/barn_og_unge/2007/bef/)." (Lesedato 27mai2008)

Vermer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. Applied Psycholinguistics 22. (2001). Cambridge University Press.

Walsh. B & Betz. N (1995) Tests and assessment. Englewood Cliffs, N.J.: Prentice Hall



Wechsler D. (1999). Wechsler Intelligence Scale for Children - Third Edition (WISC III)  
Svensk version, Manual. Stockholm. Psykologiförlaget.

Øzerk, K. Z. (2003). Sampedagogikk: en studie av de norskspråklige og  
minoritetstospråklige elevenes læringsutbytte på småskoletrinnet i L97-skolen. Vallset,  
Oplandske bokforlag.

Øzerk, K. Z. (2007). Avvik og merknad: opplæringslovens § 2-8 og enkeltvedtakordningen i  
rundskriv Udir-tilskot-10-2005 - intensjoner og konsekvenser. Vallset, Oplandske  
Bokforlag.