

Kunnskap og vurdering

Steinar Bøyum

1. Innleiing

Emnet i dette essayet er vurdering. Men eg kan ikkje ta for meg vurdering i si fulle breidd. For det første skal eg avgrense meg til summativ vurdering. For det andre skal eg konsentrere meg om strukturen i ein filosofisk debatt. For det tredje skal eg snakke om epistemologi snarare enn etikk og politikk. Likevel har artikkelen eit breiare siktemål enn kva dette kan tyde på. Strukturen i den epistemologiske debatten eg skal ta opp er nemleg eksemplarisk. Indirekte tek eg difor sikte på å seie noko meir generelt om høvet mellom filosofi og pedagogikk, mellom teori og praksis. Eg trur dette er relevant for måten filosofiske teoriar ofte blir handtert på innan mykje pedagogikk.

Vurdering er innvove i konflikt. Prøvar, karakterar, eksamen, rangering og måling er kontroversielle ting. Til dels er det slik fordi feltet er politisk lada. Ein skal trø varsamt her for ikkje å setje sinn i kok. Diskusjonen i dette essayet tek sikte på å vere uavhengig av spesifikke politiske syn, sjølv om nokre sikkert vil meine at det ikkje har lukkast meg. Eg vil likevel hevde at dei problema eg skal ta for meg gjeld uavhengig av politisk ståstad, om enn måten ein reagerer på desse problema på er avhengig av politisk syn.

Nokre av problema kring vurdering er praktiske. Korleis kan ein til dømes konstruere betre prøvar? Andre spørsmål er empiriske. I kva grad påverkar til dømes vurdering den måten det blir undervist på? Men det finst også filosofiske spørsmål om vurdering av læring. Nokre av desse er av etisk og politisk art. For som R. F. Dearden ein gong skreiv: "Wrongly to assess a person is not simply to make a mistake but, particularly if he is under-rated, also to do him an injustice" (1979, s. 115). Men mange av problema er epistemologiske: Kan kunnskap

målast? Korleis kan ein vite kva andre kan og veit? Kan vurdering vere sikker eller sikker nok? Det er slike spørsmål me skal krinse rundt i det følgjande.

2. Grammatiske strukturar

Når ein skriv om noko, tek ein gjerne til med ein definisjon eller to, slik at lesaren veit kva ein skriv om. Det skal eg ikkje gjere. Eg skal heller gi ei røff oversikt over sentrale konseptuelle strukturar kring vurdering, både for å leggje på plass ein terminologi og for å nærme oss problema me skal drøfte. Ein kan tenkje på det som ei skisse over vurderingsomgrepets logiske grammatikk, i Wittgensteins forstand (1953/1997). For å gjere det skal eg lene meg på ein artikkel av nettopp Dearden, som gir eit fint bilete av særleg dei epistemologiske bitane av denne grammatikken (1979, s. 111-118). Eg vonar at dette biletet er mest mogleg ukontroversielt, noko som dei fleste av oss i alle fall i utgangspunktet kan akseptere eller anerkjenne som vårt omgrep om vurdering, sjølv om jamvel dette elementære biletet kan bli gjenstand for filosofisk tvil, som me skal sjå.

Vurdering av læring siktar mot ein type kunnskap om eit anna menneske, nemleg om kva nokon veit, kan eller forstår. For at det skal vere mogleg, må det ein har lært bli uttrykt på ein eller annan måte. Det lærte må eksemplifiserast i noko som blir sagt, skreve, gjort eller laga. Dette produktet dømer ein så som sant eller falskt, riktig eller galt, godt eller dårlig. Denne dommen vil, som oftast saman med andre dommar, danne grunnlaget for ei slutning til ein meir generell dom om barnet, eleven eller studenten har erverva seg ei viss *evne*. Ei slik slutning treng ein ikkje trekkje ved til dømes estetisk vurdering: Då kan ein nøye seg med å felle ein vurderande dom om sjølvve verket, utan å vere oppteken av kva kunstnaren kan og ikkje kan. Men ved pedagogisk vurdering er me ikkje primært interessert i prestasjonen i og for seg, men kva han seier om personen sine evner generelt. Ei evne, tenkjer ein seg, viser seg i det ein presterer, men er sjølv meir permanent, og gir grunnlag for liknande prestasjonar i framtida.

Problema oppstår fordi ei slik slutning aldri er logisk tvingande. Det er alltid ei kløft mellom eksempelet (produktet eller prestasjonen) og den evna eksempelet eventuelt viser, uttrykkjer eller er belegg for. Men kor stor kløfta er varierer etter kva evne det er tale om. Visse fysiske eller praktiske evner er ganske eintydige. Dersom ein fint sym 50 meter, er det liten tvil om at ein kan symje. Det er logisk mogleg at det berre var flaks, men det er berre logisk, og ingen andre enn filosofar lagar problem ut av det.¹ Epistemologien er likevel ikkje symmetrisk her. For dersom ein elev *ikkje* klarar å symje 50 meter, er slutninga vanskelegare å trekkje: Kanskje var han sjuk, kanskje syntest han vatnet var for kaldt, kanskje misforstod han instruksjonane etc. Nokre av desse tinga kan ein utelukke dersom ein tek fleire prøvar, gjerne på ulike tidspunkt, eller dersom ein graderer evnene det er snakk om. Dette kan medføre praktiske problem, men ikkje prekære filosofiske problem.

Ved kognitiv eller intellektuell læring er gapet større og ting i det heile meir komplisert. Det eleven då eventuelt kan, veit eller forstår blir primært uttrykt gjennom språk, og ein kan uttrykkje det same på ulike måtar og ulike ting på same måte. To studentar kan seie det same, "Den transcendentale deduksjonen er ikkje gyldig", men i det eine tilfellet kan det vere uttrykk for kunnskap (eller forståing) og i det andre tilfellet ikkje. Og omvendt kan same kunnskap bli uttrykt på utallege måtar, og som dessutan ikkje let seg avgrense klart på førehand: Ein kan vise si forståing for Kant både i eit enkelt bilete og i ei lang avhandling. Dette gjer slutninga frå eksempel til evne meir hasardiøs enn kva tilfellet er med enkle fysiske evner. På toppen av dette kan han som skal bli vurdert ha ein dårleg dag, vere nervøs, misforstå, ha flaks etc., og han som skal vurdere kan også ha ein dårleg dag, vere trøtt, misforstå, vere inkompetent etc. Det er nærast inga ende på slike kjelder til feilvurdering, kva Dearden kallar "inferential hazards", farar ved å slutte frå eksempel til evne i vurdering av læring (1979, s. 114).

¹ Det er naturlegvis meir enn berre logisk mogleg at ein kan miste evna til å symje, men det gjer ikkje den opphavlege vurderinga ugyldig – tvert imot, at den var gyldig er ein føresetnad for å kunne tale om å *miste* evna.

Dei praktiske problema kring vurdering er særleg retta mot å redusere slike slutningsfarar. Ein kan hente inn eit større og meir variert utval av prestasjonar, ein kan ha fleire og meir øvde sensorar, ein kan formulere klarare kva som skal bli vurdert og på kva grunnlag etc. Men kan ein fjerne risikoen for feilvurdering? Kan ein vite sikkert om nokon har lært noko eller ikkje? Dersom ein med "sikker" meiner logisk eller matematisk sikker, så kan ein ikkje det, i følge Dearden (1979, s. 115). Ein kan minke sjansen for feilvurdering, men ikkje nulle han ut. Det vil alltid vere ei kløft mellom eksempel og evne. Det tyder ikkje at vurderinga ikkje kan vere sikker *nok*, gitt andre kontekstuelle trekk, særleg formålet med vurderinga. Då melder det seg eit vell av praktiske omsyn, samt etiske (rettstryggleiken til han som skal bli vurdert) og økonomiske (ressursar ein nyttar på vurderinga). Her har ein ikkje noko anna val enn å balansere dei epistemologiske omsyna med dei etiske, politiske og økonomiske og å vurdere frå sak til sak kva som er formålstenleg.

Me nærmar oss filosofiske minefelt no, men før me hoppar ut i dei må me også nemne korleis epistemologien blir endå meir kompleks når me reknar inn ulike typar vurdering. Vurdering kan vere formell (skjer i ein spesiell kontekst styrt av eigne reglar og roller og der desse er kjende for deltakarane) eller uformell (skjer som ledd i livets vanlege flyt, kanskje utan at den vurderte veit om det), formativ (som har til hensikt å fremje læring) eller summativ (som har til hensikt å lese av kva eleven har lært), prøvebasert (går føre seg i eit avgrensa tidsrom, som respons på ei oppgåve) eller observasjonsbasert (går føre seg kontinuerleg med andre aktivitetar, ikkje som svar på oppgåve).

Desse distinksjonane er ikkje parallelle. Alle slags kombinasjonar er tenkjelege. Den tradisjonelle eksamen er formell, prøvebasert og for det meste summativ, men vurdering av læring er eit mykje vidare fenomen enn som så. Når eg står og ser på son min som prøver å trikse med ein ball, er vurderinga mi uformell og observasjonsbasert. Dersom eg ber han om å passe ballen til meg, for å sjå kor godt han treff, er det ei uformell, men prøvebasert vurdering. Og i begge tilfeller kan vurderinga vere enten formativ eller summativ, avhengig av om og korleis eg deler vurderinga mi med han. Dersom dei så gjer den same prøven på fotballaget

som ledd i å ta gullmerket, er vurderinga formell og summativ, men gjerne også formativ. I det følgjande skal me ta for oss summativ vurdering, men det utelukkar altså ikkje at poenga også angår til dømes uformell og formativ vurdering.

3. Skeptisk kritikk

3.1. Rik kunnskap

I ei rekkje kontroversielle artiklar, særleg frå 1998, leverer Andrew Davis eit sofistikert filosofisk angrep på vurderingssystema i moderne utdanning (1998h) Slik vil han gje filosofisk substans til den vage kjensla mange har, at utdanning "ikkje kan målast", og dermed omsetje den vage kjensla til aktiv motstand mot testregimet i staden for passiv aksept av det. Desse skeptiske argumenta skal me no sjå nærare på. Dei undergrev tilsynelatande det elementære biletet av vurdering som me presenterte ut frå Dearden.

Davis leverer mange og varierte argument mot testing og vurdering, også praktiske og empiriske. Eg skal her konsentrere meg om dei filosofiske, sjølv om eg, for å føregripe litt, kjem til å konkludere med at dei praktiske og empiriske argumenta likevel er dei viktigaste.² Men det kunne me ikkje vite utan ei slik undersøking, for filosofiske problem har den karakteren at dei liksom *overveldar* andre problem. Praktiske og empiriske problem er liksom så *ubetydelege* i høve til dei filosofiske. Slik tenkjer også Davis: Å overvinne praktiske og empiriske problem er ikkje nok – sjølv den perfekte vurdering er filosofisk mangelfull (1998d, s. 68)

Det som gjer Davis sin kritikk interessant, er at han går med på motstandarens spelereglar. Kritikken føreset ikkje høgtflygande tankar om danning og slikt. Strategien er heller å gå med på at målet for utdanning er å bidra til ein sunn og

² Davis skriv ut frå ein engelsk kontekst, og sjølv om situasjonen der er verre (eller betre, avhengig av korleis ein ser det) enn i Noreg, så er dei filosofiske argumenta hans like gyldige i ein norsk kontekst, dersom dei då er gyldige naturlegvis.

konkurransedyktig økonomi. Argumenta hans tek så sikte på å vise at *sjølv gitt dette målet* er det djupe prinsipielle problem med prøving og vurdering, slik at det heller ikkje er i økonomien si interesse å oppretthalde "testregimet".

Personleg meiner rett nok Davis at utdanning er og bør vere meir enn ein tenar for konkurranseevna, men poenget er at argumentet hans ikkje kviler på det (1998a, s. 24).

Kva er det så den moderne økonomien treng? Kva treng arbeidstakarar for å vere til nytte i arbeidslivet? Davis svarar: Rik kunnskap (1998a, s. 26-38.). Det tyder for det første kunnskap som ein *forstår*. Å forstå noko betyr å kunne setje det i ein samanheng, forbinde det på ein passande måte med andre ting ein veit, kan og forstår. Som oftast inneber det også ei viss forståing for grunnane som kunnskapen kviler på. Dette skil rik kunnskap frå tynn, isolert eller mekanisk kunnskap. Heller ikkje økonomien har nytte av kunnskap som ein lirer av seg på papegøyevis, om då noko slikt i det heile kan kallast for kunnskap.

For det andre er rik kunnskap ein form for kunnskap som ein kan *bruke*. Ein må kunne anvende kunnskapen også i andre kontekstar enn dei som ein tileigna seg han i, jamvel i andre enn skuleliknande samanhengar i det heile. Også økonomien treng kunnskap som ein kan dra nytte av og kan anvende fleksibelt i ulike kontekstar. Dette tyder ikkje at rik kunnskap er det same som praktisk kunnskap. Rik kunnskap er kunnskap som kan nyttast i eit mangfald av kontekstar, både teoretiske og praktiske, og verken reint teoretisk eller reint praktisk kunnskap er av særleg verdi i dagens samfunn, der kravet om at kunnskap kan nyttast går hand i hand med kravet om at ein skal kunne argumentere for og reflektere over det ein seier og gjer. Slik er rik kunnskap snarare ein syntese av teoretisk og praktisk kunnskap.

Premissen for Davis sitt argument, og for meg i det følgjande, er altså at utdanning dreier seg om at elevar og studentar tileignar seg slik rik kunnskap. Og den sentrale påstanden Davis så set fram, er at *rik kunnskap ikkje let seg vurdere* (t.d. 1998d). Om og i kva grad nokon har rik kunnskap på eit felt let seg ikkje lese av frå eit avgrensa eksempelsett. Rik kunnskap manglar den spesifikke

karakter som gjer at han kan målast. Difor eksisterer rett og slett ikkje utdanningsutbytte i den finkorna forstand som moderne vurderingssystem føreset. Berre ved vurdering av svært smal eller tynn kunnskap, enten det er av teoretisk eller praktisk art, er det mogleg å kombinere høg reliabilitet og validitet, og slike evner er det heller ikkje i arbeids- og næringslivet si interesse å framelske. Davis intenderer dette som ein konseptuell kritikk. Argumenta er meint å følgje av sjølv *omgrepet* rik kunnskap. Dersom det er riktig, tyder det at betre vurderingssystem heller ikkje kan hjelpe. Sjølv prosjektet er feilslått.

Me skal no sjå litt på Davis sine argument for denne posisjonen. For ordens skuld skal eg dele dei i to, tilsvarande skiljet mellom teoretisk og praktisk kunnskap. Som nemnd heng desse to nært saman for Davis, sidan rik kunnskap er ei syntese av teoretisk kunnskap (tru) og praktisk kunnskap (evner) og sidan begge argumenta kviler på dei same grunnpoenga om likskap og disposisjonar. Denne inndelinga er difor berre eit analytisk grep for å skape orden i framstillinga.

3.2. Det teoretiske aspektet

Det teoretiske aspektet tyder at rik kunnskap alltid har, om enn i varierende grad, eit element av deklarativ kunnskap. Slik påstandskunnskap blir ofte definert som grunngeven sann tru, og sjølv om denne definisjonen er filosofisk kontroversiell, så rekk han for oss her. I alle fall verkar det trivielt at påstandskunnskap involverer det å halde visse påstandar for sanne. Å vite at Svartedauden kom til Noreg i 1349 er å tru at Svartedauden kom til Noreg i 1349, at denne trua er sann og eventuelt å kunne grunngeve ho, eller i det minste å vite kvar grunnane er å finne. Davis går difor vegen om omgrepet tru for å argumentere for sitt syn (1998b; 1998c).

Davis byggjer på ei familie av filosofiske syn som me kan samanfatte som *holisme*. Lat oss byrje med eigenskapsholisme. Kva det er å vere ein bank heng saman med kva det er å vere pengar, lån, innskot etc. Ein viss eigenskap er kva han er gjennom relasjonar til andre eigenskapar. Eigenskapar eksisterer berre i klyngjer. Dette leiar til eller heng saman med holisme med omsyn til påstandar.

Dersom det å vere ein bank er uløyseleg forbunde med det å vere pengar, er ikkje påstandar om bankar uavhengige av påstandar om pengar. Påstandar eller proposisjonar er ikkje diskrete og spesifikt identifiserbare "ting". Dei eksisterer også berre i klyngjer. Vidare gjeld det same for påstandshaldningar, til dømes det å tru noko, sidan å tru er å halde ein påstand for sann. Davis:

I cannot believe that plants require water to live and grow without beliefs about what it is to be a plant, the nature of growth, life and water. In turn these beliefs bring others in their train. Beliefs do not exist as single items. Speaking of single beliefs is a way of gesturing at an aspect of a whole network of beliefs whose individual elements cannot fully be identified independently from one another. (1995, s. 6)

Når me tilskriv nokon ein bestemt oppfatning, så tilskriv me ho samtidig eit komplekst nettverk av andre oppfatningar. Sagt på ein annan måte, for å tilskrive nokon ein bestemt oppfatning, må me også ha omfattande kjennskap til kva anna ho trur. Dommen om at nokon trur ein viss ting er ikkje ein påstand om at det finst ein viss ting inni ho. Truer er ikkje sjølvstendig eksisterande. Sidan tru er den subjektive sida av kunnskap, gjeld det same det å ha kunnskap, i det minste den kunnskapen som utdanning har som mål, rik kunnskap.

Men holismen famnar endå vidare for Davis. Ei tru er ikkje berre konstituert av relasjonar til andre truer, men også av sin plass i eit vidare felt av interesser og intensjonar, og alle desse er igjen konstituert av handlingar. Å ha ei viss tru er, i det minste mellom anna, å vere disponert for å handle på visse måtar. "Tru" er meir som "bord" enn "vatn", definert av funksjon snarare enn struktur (mykje kan nyttast som bord, men berre H₂O kan vere vatn) (Davis, 1998c, s. 64.).

No trengde ikkje holisme vore noko stort problem for vurdering, dersom det var klare, enkle og eintydige kriterium for ulike truer. For å få greie på om nokon visste ein viss ting kunne ein då berre krysse av på ei liste med kriteria for det å vite den tingen. Diverre finst det ikkje slike lister. Truer er nemleg det ein kallar "fleirspora" disposisjonar: Ulike spor kan føre til same tru, og like spor kan føre til ulike truer (Davis, 1998c, s. 62-63). Ja, gitt andre truer, ønske og intensjonar kan nesten einkvar handling manifestere ei viss tru. Å tru at det kjem til å regne kan vise seg både i at ein tek med seg paraply og at ein ikkje tek med seg paraply

(kanskje ein har lyst å bli våt), både i at ein tek på seg regnfrakk og at ein tek på seg ei jakke som ikkje er vasstett (kanskje ein trur ho er vasstett). Dette viser kor komplisert det kan vere å tilskrive ei bestemt tru til ein annan person, og denne kompleksiteten gjer at det same mønsteret av ønske, intensjonar og handlingar kan tolkast på ulike måtar som alle passar like godt med det gitte mønsteret. Det betyr at om nokon har ei viss tru eller ikkje, og dermed om nokon veit noko eller ikkje, er radikalt ubestemt og ikkje objektivt gitt.

Kva tyder dette for vurdering? Det undergrev i følgje Davis det biletet av vurdering som ligg under trua på prøvar og testar innan utdanning. I dette biletet dreier vurdering seg om å få kunnskap om det kognitive møblementet i elevanes sinn, å finne ut kva bitar av kunnskap som er der inne og kva som ikkje er der. Mot dette stiller Davis opp eit anna bilete. Kva nokon trur, og dermed kva kunnskap nokon har, er resultat av ein kompleks og kontekstsensitiv interpretasjon av det dei seier og gjer, der me også trekkjer inn andre truer, ønske og intensjonar, samt andre ting dei har sagt og gjort, ja, i yttarste forstand heile deira livshistorie. Det tyder for det første at for å ta stilling til om nokon veit noko, så må me allereie ha brei og djup kjennskap til den personen det gjeld, noko som gjer standardiserte prøvar inadekvate.³ Vidare tyder det at det vil vere stor usemje i slike tolkingar, og at det er lite me kan gjere med det, noko som betyr at høg inter-rater reliabilitet er uoppnåeleg. Og for det tredje tyder Davis' si antirealistiske oppfatning at dersom me spør om nokon har ei viss tru på eit visst tidspunkt, så finst det ikkje noko absolutt sant eller falskt svar på det. Sidan kunnskap er ein type tru, så impliserer det at det å tilskrive kunnskap til andre er noko langt lausare enn det testregimet innan dagens utdanning føreset. Ja, det synest å bety at vurdering av læring innan utdanning strengt tatt ikkje er mogleg i det heile tatt, fordi me ikkje kan oppnå fast og påliteleg kunnskap om kva kunnskap andre har.⁴

³ Dette er noko White (1999, s. 203-204) legg stor vekt på i eit forsvarsskrift for Davis.

⁴ Hugs at me heile vegen her framleis snakkar om *rik* kunnskap.

3.3. Det praktiske aspektet

Me går no til det praktiske aspektet ved rik kunnskap (Davis, 1998e). Vurdering verkar å føresetje eit omgrep om evne, eventuelt nærskyldte omgrep som eigenskap eller kompetanse.⁵ Når me vurderer, dreg me ei slutning frå eit avgrensa sett eksempelprestasjonar til den evna eksempla eventuelt manifesterer, til dømes lese, subtrahere og tenkje kritisk. Evneomgrepet er basert på eit omgrep om overføring, som igjen er basert på eit omgrep om likskap. Å ha tileigna seg ei evne er å bli i stand til å prestere *det same* i ei rekkje andre men likevel *same type* kontekstar, til dømes å løyse same type oppgåve på same måte. I dette argumentet mot vurdering kritiserer Davis vrangførestillingar om evner nettopp gjennom ei drøfting av omgrepet likskap.

På overflata verkar menneskelege evner å vere parallelle med fysiske "evner". Vatn har visse evner/eigenskapar (koke ved 100): Det oppfører seg på same måte (kokar) i same type situasjon (oppvarming til 100). Ein person har også visse evner/eigenskapar (sjananse, kan lese, meistarar addisjon): Han oppfører seg på same måte (blir taus, les, svarar riktig) i same type situasjon (i selskap, skal lese noko, får ei addisjonsoppgåve). Men denne parallellen stikk ikkje særleg djupt. Kjemiske stoff har underliggjande strukturar som forklarar stoffets evner og som gjev substans til identifikasjon av dei. Kva som ovanfor gjeld som "det same" kan utmyntast i fysiske termar, atomære og molekylære. Men slik er det ikkje for menneskelege handlingar og evner. Kva som då gjeld som det same er ikkje objektivt gitt. Når me spør, stilt ovanfor ein viss prestasjon eller handling, om han er uttrykk for den same eller ei anna evne, finst det ikkje noko gitt svar. Til sjuande og sist kan me velje om me vil rekne noko som det same eller ikkje. Kva me vel blir bestemt av hensikter, interesser og kontekstar. Dermed gjeld det same også evner. Evner, og klassifikasjon av desse, er ikkje objektivt gitte på same måte som fysiske eigenskapar og disposisjonar.

Her kviler Davis på Nelson Goodman på den eine sida, og Jean Lave og Etienne Wenger på den andre (1998e, s. 83-86). Frå Goodman hentar han tanken om at

⁵ Eg ser her vekk frå skilnadar mellom desse omgrepa, og held meg frå no av til evner. Eg unngår også diskusjonar om (teoretiske) "constructs" meir generelt.

alt liknar på alt i ein eller annan forstand, men at kva me reknar som *relevante* likskapar i ein bestemt situasjon avheng av partikulære ting i den situasjonen: våre hensikter, interesser etc. Frå Lave og Wenger hentar han tanken om at handlingar ikkje eksisterer i isolasjon, men berre har meining i sosiale kontekstar og fellesskap. Det betyr, sluttar Davis, at klassifikasjon av menneskelege handlingar, inkludert tale- og skrifthandlingar, er kontekstuel bestemt. Sagt på ein annan måte, kva som gjeld som "det same", til dømes "same prestasjon" og "same oppgåve", er sosialt konstruert. Og dermed er også evner det. Om det er evne x eller y som blir utført i ein viss kontekst, om det er same type handling som blir utført i same type situasjon, har ikkje noko objektivt gitt svar.⁶

La oss sjå kort på eit av Davis sine eksempel (1998e, s. 78-94). Me reknar "4-2", "8-3" og "3-8" som same type reknestykke, nemleg subtraksjon, og me seier at den som kan svare riktig på slike og liknande oppgåver har tileigna seg ei evne til å subtrahere. Det er det sosialt konstruerte settet av omgrep som me kallar matematikk som avgjer kva me reknar som likskap og skilnad. Ved å lære matematikk så lærer barnet seg at "3-8" er *av same type som* "8-3" og tileignar seg dei likskapsomgrepa som det matematiske samfunnet har. Men det er ikkje gitt på førehand at "8-3" er det same som "3-8". Det er trass alt viktige skilnadar også, og i andre kontekstar, andre kontrafaktiske historiske prosessar, kunne me rekna skilnadene som viktigare enn likskapane. Det følgjer at det er prinsipielt umogleg å identifisere noko bestemt i barnet, ein subtraksjonskompetanse, som ligg under og sikrar riktige svar på det me reknar som same type oppgåver.

Konklusjonen til Davis er at snakk om evner er "mytologisk" (1998e, s. 106). Det finst strengt tatt ikkje noko slikt. Det ligg ikkje noko objektivt under våre handlingar, til dømes i sinnet, som styrer oss og sikrar suksess i eit mangfald av kontekstar når det fyrst har slege rot. I beste fall kan me snakke om evner som ei røff oppsummering av kva ein person til no har hatt ein tendens til å gjere. Med omsyn til eksempelet over kan me kan tale laust om eleven sin suksess på oppgåver som me er tilbøyelege til å kalle subtraksjonsoppgåver, og kanskje i lys

⁶ Me ser som vanleg vekk frå reint fysisk-praktiske evner som til dømes ballkasting.

av dette ha ein vag optimisme om korleis ho vil gjere det i framtida, men meir presist enn dette kan me ikkje gjere det (s. 94). Problemet er at vurdering innan utdanning, særleg høginnnsats vurdering, føreset eit langt meir robust omgrep om evner eller kompetansar. Då føreset ein at evner og kompetansar kan la seg avlese frå eit avgrensa eksempelsett på ein rimeleg presis og objektiv måte. Denne føresetnaden er ikkje oppfylt, og testregimet kviler difor ikkje på eit solid epistemologisk grunnlag.

4. Logisk-retoriske mønster

Det er mykje ein kan kritisere Davis for, og det er det fleire som har gjort, ofte på ein god måte (Curren 2004; Winch & Gingell, 1996; Winch & Gingell, 2000). Eg skal her berre ta opp noko som er av interesse ut over vurderingsdebatten, og som kanskje gjeld pedagogikkfaget si handsaming av høvet mellom (filosofisk) teori og (pedagogisk) praksis meir generelt.

Davis sin filosofiske analyse er intendert å underminere testregimet i dagens utdanningsinstitusjonar. Det er ein intervensjon i samtidas utdanningsdebatt. Men det han legg fram er ein generell filosofisk posisjon, som han ein stad kallar radikalt skeptisk (Davis, 1998d, s. 68). Og då er spørsmålet om han ikkje skyt sporv med kanoner. *All* vurdering blir umogleg dersom me aksepterer Davis sine poeng. Men då er det viktig å minne seg om at vurdering av læring er eit heilt kvardagsleg fenomen. Det er ikkje berre i testsamanhangar at ein seier ting som "Han kan lese", "Ho veit namnet på alle Nordens hovudstader" eller "Han er flink å rekne". Det kan difor sjå ut som om Davis kastar barnet ut med badevatnet.

Ingen holist vil få oss til å slutte å identifisere individuelle oppfatningar og bitar av kunnskap. Dei vil ikkje ha oss til å late vere i praksis å seie ting som "Han trur x" eller "Ho veit y". Det dei seier er at ei enkelt tru blir konstituert av relasjonar til andre truer og difor ikkje kan eksistere uavhengig av dei. Dette er ei teoretisk innsikt som i seg sjølv ikkje har praktiske følgjer. Men *dersom* holismen til Davis skal undergrave vurdering i utdanning, noko han meiner han gjer, og altså føre til

endring i praksis, så vil han undergrave *einkvar* identifikasjon av individuelle truer. Og det er verkeleg eit ekstremt synspunkt. For då undergrev han eit gjennomgåande trekk i vår måte å snakke om andre menneske på også utanfor formelle utdanningsinstitusjonar.

Også Davis sin posisjon med omsyn til evneomgrepet, er meir radikalt, eventuelt ekstremt, enn kva Davis verkar å vere klar over. I følgje Davis er det ikkje noko slikt som evner, *i prinsippet* (1998g, s. 94). Og nok ein gong, dette kan vere ei interessant teoretisk innsikt, men konsekvensen for praksis er langt frå klar. Men uansett, *dersom* denne "strengt tatt" skal ha praktiske følgjer, noko Davis meiner den har, dersom den skal få oss til å endre vurderingspraksis innan utdanning, så rammar dette også vurdering i det daglege. Dersom formell summativ vurdering innan dagens utdanning fell som følgje av Davis sine argument, fell også den daglegdagse vurderinga av evner og eigenskapar. For den analysen av evneomgrepet som Davis gir, gjeld evner generelt, og ikkje berre evner slik dei figurerer i psykometriske utdanningsdiskursar.

På bakgrunn av dette er det ikkje rart at Davis andre stader trekkjer seg tilbake til meir moderate posisjonar. Når han er i dette meir moderate hjørnet, får tesane hans ein ganske annan lyd. Då nøyer han seg med å hevde at det *i prinsippet* ikkje er nokon form for vurdering som kan avlese rik kunnskap *presist, i detalj* eller *med sikkerheit* (t.d. 1998d, s. 67). Og med omsyn til evner går han med på at evner kan vere "plastiske", dvs. at me kan predikere på basis av suksess i ein gitt kontekst at eleven også vil lukkast i andre kontekstar (1998e, s. 89-90). Elevar som lukkast med ein type subtraksjon tenderer å lukkast på andre subtraksjonar. Men han legg til: Ikkje alltid, og kanskje heller ikkje så ofte som me trur. Det skjer ofte at folk ikkje handlar på "same" måte i "liknande" kontekstar. Små skilnader i oppgåve eller samanheng kan til dømes føre til store skilnader i prestasjon. Evner er ikkje alltid pålitelege.

På bakgrunn av slike modereringar blir det også tydeleg at han likevel ikkje meiner at vurdering innan utdanning er umogleg: *Me kan* felle fornuftige dommar om kva elevar veit, forstår og kan, men desse dommane er røffe og

breipensla (1998f, s. 108). Då blir det også klart at han støtter ein viss type vurderingssystem (1998g, s. 151-152). Han ønskjer å ta opp igjen det som ein hadde i England rundt 1980, tilsynelatande ikkje ulikt Noreg sine nasjonale prøvar: Eit stort tal prøvar på visse alderssteg for å undersøkje nivå og prestasjonar, men utan namngjeving av elevar, klassar og skular, og der målet er kunnskap, ikkje rangering.

Kva er så problemet med denne moderate forma? Ingenting, i alle fall ikkje filosofisk. Og det er nettopp det som er problemet. Ho er ikkje berre plausibel, ho er *for* plausibel. Ho har lita eller inga injurierande kraft, sjølv den argaste forsvarar av testing i skulen kan gå med på dette utan fare for sin eigen posisjon. Me veit at vurdering sjeldan er fullstendig presis og sikker, og me veit at evner ikkje garanterer suksess – å hevde såpass kan knapt kallast ei djuptgåande kritikk av dagens vurderingssystem innan utdanning. Sant nok, dette avheng av akkurat kva Davis meiner med "presis", "sikker" og "garantere". Problemet er at Davis med slike termar ikkje verkar å meine stort anna enn "immun mot feil". Så dermed kan det sjå ut som om den sentrale påstanden hans står i fare for å koke ned til at me aldri kan vere hundre prosent sikre på at me vurderer riktig. Men det visste me jo frå før.

Faktisk kan ein undre seg over om det er Davis sjølv som ligg under for ein illusjon om vurdering. Han verkar å ha ein idé om at vurdering må vere fullstendig presis, eintydig og sikker for at vurdering i skule og utdanning skal kunne late seg gjere. Men har me eigentleg nokon idé om kva ein fullstendig presis, eintydig og sikker vurdering av rik kunnskap skulle vere for noko? Kan me tenkje oss at dommar av typen "Han kan analysere dikt", "Han er flink å rekne" eller "Han veit mykje om vikingane" skulle kunne vere like presise, eintydige og sikre som matematiske utsegn? Kunne desse dommane då betydd det same som i dag? Ville det då vore *vurdering*? Dersom me prøver å tenkje oss ei verd der dette var tilfellet, vil me ikkje få det til. Det er eit fantasifoster.

Men snarare enn å vere eit argument mot vurdering, er dette eit argument mot argumentet til Davis. Davis trur han nektar for noko, men kva dette *noko* er, er

uklart, og dermed er det Davis si eiga tese som er uklar. Så det underlege er, og dette er sannsynlegvis det stikk motsette av Davis sine (gode) intensjonar, at det er *han* som har ein metafysisk fantasi om kva vurdering er og krev, ikkje dei som forsvarar den summative vurderinga si rolle i utdanninga, til dømes Christopher Winch og John Gingell (1996, s. 380-382). Det er dei som har den rimelege forståinga av fenomenet det dreier seg om: Vurdering er ein daglegdags aktivitet, både i og utanfor skulen, som ikkje krev perfekt presisjon. Å tru at vurdering krev perfekt presisjon er å forfalske fenomenet frå byrjinga av, og å avsløre eins eigne forfalskingar avslører ikkje andre enn ein sjølv.

På mange måtar følgjer Davis sin argumentasjon eit vanleg mønster i filosofien, som me særleg kjenner frå diskusjonar om skeptisisme: Det er ei kløft mellom noko me gjer til vanleg (vurdering av læring eller andre kunnskapskrav) og det som me (etter skeptikarens oppfatning) siktar mot når me gjer dette (idealet), og denne kløfta let seg ikkje bruleggje (skeptisisme). Problemet er då at det typisk er skeptikaren som opererer med eit fantasibilete av det me gjer til vanleg, altså har det idealiserte bilete av den daglegdagse aktiviteten, som han så, logisk nok, hevdar er uoppnåeleg. Ein kan seie det slik. Målt mot den perfekte vurdering er einkvar faktisk vurdering røff og omtrentleg, men det er ingen mangel ved faktisk vurdering, men eit symptom på at me måler røynda mot ein fantasi.

Ein liknande svakheit finn me i Davis sitt tilbaketog med omsyn til evner. Frå å argumentere for at det eigentleg ikkje finst noko slikt som evner, modererer han seg til å seie at me kan bruke evneomgrepet til å oppsummere tidlegare prestasjonar, og modererer seg endå meir til også å innrømme at me kan føresjå på basis av suksess i ein gitt aktivitet at ein også vil lukkast med liknande aktivitetar i framtida (t.d. 1996, s. 302). Kritikken har då krympa inn til ein empirisk påstand om at ein ofte ikkje handlar på same måte i tilsynelatande liknande kontekstar, til dømes korleis små skilnader i oppgåveformulering kan føre til store skilnader i prestasjon. Men det er jo også mykje empirisk belegg for at ein ofte handlar på same måte i liknande kontekstar: Når ein har lært å lese, så kan ein normalt lese mange forskjellige typar tekstar, ikkje berre dei som ein fyrst lærte seg å lese. Men dette er eit empirisk spørsmål og let seg ikkje avgjere

på eit filosofisk nivå. Spørsmålet er då kvar det vart av den filosofiske kritikken. Davis si insistering på at hans kritikk er filosofisk ("konseptuell") får ein difor til å undre på om han hypotiserer empirisk mangfald til filosofiske manglar, og praktiske utfordringar til filosofiske problem.

Det som avteiknar seg her hjå Davis, er ein typisk teoretisk koreografi. Ein set ut med ein storslagen tese, som verkar både radikal og djup. Radikal, i det han skil seg frå det mange tek for gitt, samtidig som han gjerne maktar å appellere til vage kjensler mange av oss har. Djup, i det han kviler på filosofiske argument som verkar å bore ned til det djupaste fundament. Men så er det som om forfattaren anar at den radikale tesen også har ei rekkje uheldige konsekvensar. Han undergrev ikkje berre det forfattaren vil undergrave, men svært mykje anna også, ting som verkar uangriplege. Forfattaren merkar liksom dette og trekkjer seg tilbake. No meiner han det ikkje så radikalt likevel. Nei, eigentleg meinte han berre å ... og ender opp med ein langt meir moderat posisjon, som gjerne er meir akseptabel sidan han ikkje har dei uheldige implikasjonane som den radikale utgåva hadde. Ulempa er at den moderate påstanden heller ikkje er særleg oppsiktsvekkande. Han seier ikkje meir enn det me allereie visste og meinte. I verste fall er han triviell.

Problemet er ikkje berre at ein vekslar mellom ein radikal og ein moderat tese. Problemet er at ein ikkje er klar over at ein gjer det. Ein ser ikkje at ein glir frå det eine til det andre, men trur ein er på same stad heile tida. I røynda sjonglerer ein med to ballar, ikkje ein, og desperat prøver ein å halde begge ballane i lufta samstundes. Ein prøver å smelte to posisjonar saman i ein posisjon som er både djup-radikal og plausibel-akseptabel. Ein svevar i limbo mellom det radikale og det moderate, utan å setje seg ned, for då ville det også bli klart at det ein enda opp med var enten urimeleg eller trivielt, lik ein politisk balansekunstnar som vil tekkje to motstridande grupper på ei og same tid, og difor søker å formulere seg slik at det appellerer til begge retningar, heilt til venstre og høgre har smelta saman og ein ikkje lenger ser skilnaden.

Dette er den akademiske utgåva av kva John Holbo har kalla "the two-step of terrific triviality":

... a standard rhetorical move which has no accepted name, but which really needs one. I call it 'the two-step of terrific triviality'. Say something that is ambiguous between something so strong it is absurd and so weak that it would be absurd even to mention it. When attacked, hop from foot to foot as necessary, keeping a serious expression on your face. With luck, you will be able to generate the mistaken impression that you haven't been knocked flat, by rights. As a result, the thing that you said which was absurdly strong will appear to have some obscure grain of truth in it. Even though you have provided no reason to think so. (Holbo, 2007).

Det kan verke urettferdig mot Davis å seie at argumenta hans står i fare for å koke ned til ei vag åtvaring om velkjende farar og risker ved vurdering av læring. Faktum er nemleg er at han leverer ei bestemt oppfatning av kunnskap som ein kan tenkje seg står urørt av min kritikk, nemleg den sosialkonstruktivistiske. Det er ikkje mi oppgåve her å ta stilling til denne doktrinen. Mi hensikt har snarare vore å diskutere det kritiske siktemålet med dei filosofiske argumenta hans, altså dei praktiske implikasjonane han tenkjer seg at det sosialkonstruktivistiske biletet har. Og eg vil då hevde at Davis har eit altfor enkelt bilete av relasjonen mellom (filosofisk) teori og (pedagogisk) praksis, og at dette er noko han deler med litt for mange innan pedagogisk tenking.⁷

Ein kan seie at det finst to ulike tendensar i Davis sitt prosjekt, skeptisisme og holisme. Den skeptiske impulsen seier at det er noko me trur me kan gjere som me ikkje kan likevel. Den holistiske impulsen seier at me kan gjere alt me trur me kan gjere, men grunnlaget for det me gjer er ikkje det me trur. I utgangspunktet verkar Davis å ha den første i tankane, og det er denne som ligg til grunn for prosjektets kritiske karakter, brodden mot testregimet. Men som me har sett er problemet med denne at den også undergrev vanleg vurdering. Difor endar han opp med å stadig oftare gli over mot det holistiske, men då forsvinn også den kritiske karakteren. I 2006 hevdar han til dømes at han ikkje er oppteken av å underminere tilskrivningar av kunnskap, men berre å *forstå* slike tilskrivningar (2006, s. 6). Men kva blir det då med det kritiske elementet? Prosjektet har ingen

⁷ Dette er litt rart all den tid Davis har skreve innsiktsfullt om relasjonen mellom epistemologi og kurrikulum andre stader (Davis & Williams, 2003, s. 253-257).

brodd lenger. Ein kan ikkje slutte frå det faktum at alle menneske lever i universet til at astronomi bør vere det einaste skulefaget.

Kanskje Davis trur at det følgjer ein testkritisk posisjon ganske direkte frå hans sosialkonstruktivisme. Men det gjer det ikkje. Dersom me antek at all kunnskap, meining og tolking kviler mot ein sosiokulturell bakgrunn, så gjer også all vurdering det, frå dei beste vurderingspraksisar til dei verste, frå den mest daglegdagse uformelle vurdering til det mest sinnrike system av nasjonale prøvar. Men dermed kan ikkje ein slik generell teori nyttast til å skilje mellom ulike vurderingspraksisar og rette ein kritisk brodd mot nokre av dei. For å gjere det, må ein finne særtrekk ved formell høginnsats testing og argumentere mot det på det grunnlaget, men slike særtrekk finn ein ikkje gjennom ein generell teori. Det han treng er noko epistemologisk særmerkt ved høginnsats vurdering i høve til ordinær vurdering. Men det vil nok ikkje dukka opp, sidan "høginnsats" er ein karakteristikk av den sosiale og politiske funksjonen som vurderinga har, og denne funksjonen er *prima facie* irrelevant for epistemologien til vurderinga.

5. Avslutning

Me har i dette essayet sett på nokre tvilsame logisk-retoriske trekk: 1) Ein har ein generell teori om eit visst fenomen (t.d. læring), som ein så nyttar til å kritisere ein viss type av dette fenomenet (t.d. førelesinga) til fordel for ein annan type av fenomenet (t.d. gruppearbeid/seminar). 2) Ein glir fram og tilbake mellom ein radikal, urimeleg tese og ein moderat, plausibel tese, medan ein sjølv trur at ein opererer med ei enkelt tese som er både radikal og plausibel. 3) Ein trekkjer direkte praktiske konsekvensar av ein filosofisk eller teoretisk posisjon, til dømes eit sosiokulturelt syn på kunnskap, utan å sjå at dei praktiske følgjene berre følgjer gitt ei rekkje hjelpehypotesar, som ein ikkje diskuterer. 4) Ein samanliknar røynda med ein fantasi, finn ut at røynda ikkje strekk til og difor må endrast, utan å sjå at det då er ein sjølv som ligg under for ein fantasi, ikkje røynda.

Slike tankemønster er i mi ringe erfaring diverre ikkje uvanlege innan pedagogisk filosofi og teori, sjølv om eg diverre ikkje har kunne gi fleire eksempel på det her. Eg oppmodar lesaren til å sjå om han eller ho kan finne igjen dei same mønster i andre tekstar.

Og kvar tok dette oss når det gjeld vurdering? Kanskje ikkje så langt. Men det kan spare oss for omvegar i framtida. Det som ser ut som teoretiske snarvegar til å angripe, eller for den saks skuld forsvare, pedagogiske praksisar, viser seg etter mitt syn ofte å vere, ikkje berre omvegar, men villvegar. Med omsyn til vurdering ender eg då opp med den same konklusjonen som Curren, om enn me har gått ulik veg for å nå fram til han. Dei viktige spørsmåla om kva vurdering kan yte, er ikkje filosofiske spørsmål, men empiriske, praktiske og politiske:

This investigation leads me to the further conclusion that the important questions about what we can succeed in measuring with large-scale, high stakes tests are not conceptual or philosophical questions. In the absence of any cogent philosophical argument to the contrary, they are much more plausibly regarded as empirical questions, which cannot be properly resolved through arm chair analysis, socio-political critique, or speculation grounded in mere anecdotes. (Curren, 2004, s. 237)

Referansar

- Curren, R. R. (2004). Educational measurement and knowledge of other minds. *Theory and Research in Education*, 2(3), 235-253.
- Davis, A. (1995). Criterion-referenced Assessment and the Development of Knowledge and Understanding. *Journal of Philosophy of Education*, 29(1), 3-21.
- Davis, A. (1996). Who's Afraid of Assessment? Remarks on Winch and Gingell's reply. *Journal of Philosophy of Education*, 30(3), 389-400.
- Davis, A. (1998a). 2. Accountability and the Economy. *Journal of the Philosophy of Education*, 32(1), 19-39.
- Davis, A. (1998b). 3. Understanding and Holism. *Journal of the Philosophy of Education*, 32(1), 41-55.
- Davis, A. (1998c). 4. Belief and Language-based Assessment. *Journal of the Philosophy of Education*, 32(1), 57-65.
- Davis, A. (1998d). 5. Implications for Assessment. *Journal of the Philosophy of Education*, 32(1), 67-74.
- Davis, A. (1998e). 6. Transfer, Abilities and Rules. *Journal of the Philosophy of Education*, 32(1), 75-106.
- Davis, A. (1998f). 7. Matching. *Journal of the Philosophy of Education*, 32(1), 107-121.
- Davis, A. (1998g). 9. Is there a Future for Assessment and Accountability? *Journal of the Philosophy of Education*, 32(1), 145-152.
- Davis, A. (1998h). *The Limits of Educational Assessment*. Oxford: WileyBlackwell.

- Davis, A. & Williams, K. (2003). Epistemology and Curriculum. I Blake, N., Smeyers, P., Smith, R. D., & Standish, P. (red), *The Blackwell Guide to the Philosophy of Education* (s. 253-270). Oxford: Wiley-Blackwell.
- Davis, A. (2006). High Stakes Testing and the Structure of the Mind: A Reply to Randall Curren. *Journal of Philosophy of Education*, 40(1), 1-16.
- Dearden, R. F. (1979). The Assessment of Learning. *British Journal of Educational Studies*, 27(2), 111-124.
- Holbo, J. (2007). When I hear the word culture ... aw, hell with it. Lasta ned 12.8.2009 frå: <http://crookedtimber.org/2007/04/11/when-i-hear-the-word-culture-aw-hell-with-it/>
- White, J. (1999). Thinking about Assessment. *Journal of Philosophy of Education*, 33(2), 201-211.
- Winch, C. & Gingell, J. (2000). Curiouser and Curiouser: Davis, White and Assessment. *Journal of Philosophy of Education*, 34(4), 673-685.
- Winch, C., & Gingell, J. (1996). Educational Assessment: reply to Andrew Davis. *Journal of Philosophy of Education*, 30(3), 377-388.
- Wittgenstein, L. (1997). *Filosofiske Undersøkelser*. Oslo: Pax palimpsest. (Originalen publisert i 1953).