



# Nye evalueringsformer i et handlingsteoretisk perspektiv

**Vidar Gynnild**

**NTNU, Trondheim**

## **Summary**

Innovative or authentic assessment is a crucial part of the current reform of higher education in Norway. The author argues that assessment should be an integral part of the teaching and learning process as there is evidence that this will contribute significantly to the learning outcomes. The intention of the article is not to recommend specific methods of assessment, but rather to approach assessment from an action theoretical point of view. Assessment is seen as a driving force for learning, and therefore all the main learning objectives should be covered in the ways in which courses are assessed. To provide feedback the assessment activities should be spread across the semester. This will enable the student to identify what he or she has achieved satisfactorily or not. It is not necessarily the actual methods of assessment that should be changed; rather, it is the underlying philosophy and the application of any method of assessment that should be reconsidered.

## **1. Innledning**

Nye, eller alternative vurderingsformer, står høyt på agendaen i dagens reform av norsk høyere utdanning. Det gis det flere gode grunner til, men aller viktigst er trolig ønsket om å bidra til bedre og mer relevant læring. Tidligere ble det gjerne trukket et skille mellom undervisning og læring på den ene siden, og eksamen på den andre siden. Nå er dette i ferd med å bli snudd på hodet ved at vurderingsopplegg inngår som en del av læringsaktivitetene.

Hensikten med denne artikkelen er ikke å komme med konkrete råd med hensyn til valg av vurderingsopplegg i enkeltfag eller studier generelt. Utfordringen er altså ikke av typen: Skal jeg velge “mappevurdering” eller “hjemmeeksamen?” Jeg forsøker her å se bruken av alternative vurderingsformer i et handlingsteoretisk lys. Hvordan kan studieatferd og læring forstås i lys av ulike typer “rammer”, for eksempel ved kreativ bruk av vurderingsordningene?

## 2. Vurderingsformer og studieatferd

Nær halvparten av all norsk ungdom tar høyere utdanning. På femti år har tallet på studenter økt fra 5 000 til 170 000, samtidig som ønsket om lavest mulig strykprosent er opprettholdt. Den stadig bredere rekrutteringen til høyere utdanning bidrar til at utdanningsinstitusjonene søker nye veier i måten de organiserer og gjennomfører undervisning og læring på. Det som var gode løsninger for 30 år siden, er ikke nødvendigvis godt nok i dag. Godt motiverte, faglig sterke og selvdrevne studenter er langt fra noen selvfølge. I min undersøkelse av studieatferd blant sivilingeniørstudenter ved NTNU fant jeg at de i snitt arbeidet til dels betydelig mindre enn stipulert, og eksamensmotivert læring var utbredt.<sup>1</sup> Dette er i seg selv et velkjent fenomen i en rekke studier, men fram til de siste årene har relativt lite vært gjort fra institusjonenes side for å utvikle alternative vurderingsopplegg som læringsfremmende tiltak.

Enkelte vil stusse over at denne problemstillingen i det hele tatt blir reist. De fleste av oss er mer vant til å tenke at hensikten med eksamen er å kontrollere *resultatene* av et læringsarbeid, altså ikke som et aktivt virkemiddel underveis i studiet. Dårlig læring og høy strykprosent blir fra studentenes side gjerne forklart med henvisning til dårlig undervisning, mens faglærerne på sin side klager over sviktende forkunnskaper og dårlig arbeidsmoral blant studentene. De svarene vi kommer fram til, er altså i høy grad avhengig av de perspektivene vi legger til grunn for vår tenkning om fenomenet og sammenhenger på dette feltet. Felles forståelse for sentrale begrep kan være ett bidrag til å gjøre diskusjonen mer meningsfylt for oss. Hva vil det for eksempel si å *undervise*? La oss starte med en enkel definisjon: Å *undervise* betyr å legge til rette for læring. Alle problemer er imidlertid ikke løst med dette, for hva betyr det å *lære*? Finnes ulike former for læring, og hvordan kan vi legge til rette for det vi ønsker? De fleste av oss vet hvordan tradisjonell slutteksamen har en handlingsdrivende effekt. Det er denne "mekanikken" som utnyttes ved bruk av nye, eller alternative vurderingsformer. Skillet mellom undervisning og vurdering blir i noen grad brutt ned, med læring som sentralt mål.

Under mitt doktorgradsarbeid ved daværende Fakultet for marin teknikk ved NTNU lot jeg et utvalg studenter føre tidslogg for sitt studiearbeid gjennom et helt semester. Selv om en skal være forsiktig med å trekke konklusjoner på grunnlag av avgrensede studier, ga undersøkelsen grunnlag for ettertanke for de involverte faglærerne. De fleste studentene jobbet mindre enn stipulert gjennom semesteret, med unntak av de to siste ukene foran eksamen. Skippertak var

---

<sup>1</sup> Gynnild (2001).

utbredt, og mer omfattende undersøkelser viste betydelig innslag av memoreringsbasert og taktisk læring med eksamen som mål i seg selv, med relativt mindre vekt på dypere læring og forståelse av grunnleggende fenomen og faglige sammenhenger. Paradoksalt nok viste en sluttevaluering av undervisningen i kurset stor grad av studenttilfredshet. Studentene vurderte undervisningen som til dels mye bedre enn i andre fag de hadde hatt i studiet.<sup>2</sup> Eksemplet viser at det er mulig å få høy score på undervisningen, samtidig som mange av studentene jobber på ikke-intenderte måter, med ønsket om å bestå eksamen som overordnet mål.

I et faglig tettpakket studium som sivilingeniørstudiet, kommer studentene under tidspress. Det betyr at de fleste blir tvunget til strategiske valg med hensyn til tidsbruk og arbeidsmåter. Alle frivillige og obligatoriske aktiviteter i studier og fritid kaller på studentenes valg. Arbeid med studiene er spesielt knyttet til to fenomen, nemlig *sertifisering* og *læring*. Ideelt sett tenker vi oss sertifisering som uttrykk for et visst nivå av læring, altså som godkjenning av faglig kompetanse. Undersøkelser fra sivilingeniørstudiet viser imidlertid at enkelte av studentene finner "snarveier" i systemet, som setter ideen om sertifisering på faglig grunnlag i et nytt lys. La meg illustrere dette ved følgende sitat fra en sivilingeniørstudent:

For å få et godt læringsresultat, bør arbeidet fordeles noenlunde jevnt gjennom semesteret, med en viss økning av innsatsen i ukene før eksamen. For å få en god eksamenskarakter, vil det beste være som for et godt læringsresultat, eller man kan bruke veldig mye tid i perioden like foran eksamen, hvis man ikke har brukt noe særlig tid i semesteret. Jeg mener det er fullt mulig å få en god karakter, selv om innsatsen gjennom semesteret er ganske lav. Ulempen er at læringseffekten blir relativt dårlig.<sup>3</sup>

Studenten beskriver et studieopplegg som åpner for ulike typer studieatferd. I følge sitatet er det mulig å få en bra karakter selv om innsatsen gjennom semesteret har vært ganske lav. Teoretisk er det altså mulig å oppnå samme resultat med hensyn til karakter på grunnlag av vidt forskjellig studieatferd. I min undersøkelse av studieatferd i sivilingeniørstudiet ble dette beskrevet ved studentenes ulike tilnærming til egen læring. Enkelte studenter bruker mye av sin tid til å lære for eksamen som mål, mens kvaliteten av egen læring ser ut til å bety mindre. En annen gruppe studenter er i sterkere grad opptatt av læring

---

<sup>2</sup> Undersøkelsen ble gjennomført blant studenter (ca 100) i et marinteknisk grunnkurs i tredje studieår.

<sup>3</sup> Student NN, NTNU, 1997.

som et mål i seg selv, selv om det naturligvis også er ønskelig at de får en god karakter til eksamen. Også typen av oppgaver som blir gitt til eksamen, er for enkelte studenter et tema for inngående studier, slik følgende sitat viser:

Når det gjelder eksamenskulturen ... har jeg erfart at på de fleste eksamener har det vært visse oppgaver som er sikre gjengangere fra år til år. Jo mindre faget er, jo sikrere kan en være på at løsning av gamle eksamensoppgaver gir god uttelling. Men ved større fag er oppgavene av en mer ondskapsfull art, og fantasien hos faglærer ofte større. Min erfaring er at det i disse fagene er nødvendig med en grundigere forståelse.<sup>4</sup>

Under intervju med ekstern sensor etter eksamen blir følgende hjertesukk lagt for dagen:

Det går veldig mye på husk, de husker formler som de prøver å tilpasse problemer uten å forsøke å tenke gjennom fysikken og prøve å resonnerer på eksamen. Læringsprosessene har vært for dårlig, helt klart ... Hvis fundamentet da ikke er forstått, sitter studentene med et kjempeproblem når de skal opp til eksamen.<sup>5</sup>

Et utdanningssystem som premierer rask "gjennomstrømning" løper en risiko for dårlig læring og sertifisering av kandidater på falske premisser. Universitetet blir ikke bare en arena for konkurranse mellom studentene, men også mellom faglærere, institutter og fakulteter. Spesielt i valgbare fag kan "vanskelige" oppgaver og høy strykprosent gi negative signaler til framtidige studenter. Fristelsen kan være stor for å velge fag og studieretninger som krever moderat innsats, med utsikt til en sikker bestått eksamen. Modulisering av studiene og høy grad av valgfrihet mellom ulike kurs, stimulerer konkurransen mellom faglærere og institutter. Samtidig imøtekommer økt valgfrihet individuelle behov blant studentene, men kanskje er det noe ved akademisk læring som går tapt gjennom denne typen faglig fragmentering?

Dersom en ønsker å forklare kvalitativt ulike typer studieatferd, som sitatene foran illustrerer, kan dette gjøres ut fra ulike tilnærminger. En kan på den ene siden rette oppmerksomhet mot særtrekk ved enkeltindivider, eller studieatferd kan også tolkes i lys av strukturelle fellestrekk som studentene utøver sitt studiearbeid under. Begge perspektivene er interessante, men med ulik begrunnelse. I denne artikkelen velger vi en tilnærming der studieatferd forstås i lys av felles, ytre betingelser for handling. Jeg er altså ikke primært opptatt av

---

<sup>4</sup> Student NN, NTNU, 1997.

<sup>5</sup> Personlig meddelelse fra Dr. ing. Jan Roger Hoff, MARINTEK, Trondheim.

individuelle trekk ved studentene, men ved ulike “rammer” som fremmer bestemte typer studieatferd og læring.

### 3. Hvorfor nye vurderingsformer?

Noen vil hevde at eksamen har vist seg å være en levedyktig vurderingsform. Den har lange tradisjoner, og er relativt enkel å administrere. Hvorfor trenger vi da nye vurderingsformer? I 1997 laget Organisasjonen for økonomisk samarbeid og utvikling (OECD) i samarbeid med Kirke- utdannings- og forskningsdepartementet (KUF) en rapport om høyere utdanning i Norge.<sup>6</sup> Her blir det konkludert med at for store ressurser blir brukt på formell eksamen, mens mindre vekt blir lagt på læring og utvikling. Det blir gitt for lite tilbakemelding til studentene underveis i læringsforløpet. Alt i alt er utdanningene preget av for mye måling og for lite utvikling. “This has many features of a certification system where students tend to work only for exams ...”<sup>7</sup> Søkelys mot vurderingsformene er i høy grad også skapt under innflytelse fra utlandet. Følgende sitat er hentet fra University of Queensland i Australia:<sup>8</sup>

The primary focus of assessment is to encourage, direct and reinforce learning. Assessment should be designed to assist students in their learning.

Assessment should also be capable of indicating achievement, maintaining standards and provide certification.

The assessment system should be as transparent as possible.

Som vi ser, er hovedhensikten med vurdering å styrke studentenes læring, mens kontroll og sertifisering kommer på de neste plassene. I den norske universitets- og høyskoleloven har imidlertid denne nyansen falt bort. Det prinsipielt nye er her at det skal være ekstern evaluering av vurderingen eller av vurderingsordningene. I paragraf 50, nr. 1 står følgende:

Styret skal sørge for at studentenes kunnskaper og ferdigheter blir prøvet og vurdert på en faglig betryggende måte. Vurderingen skal også sikre det faglige nivå ved vedkommende studium. Det skal være ekstern evaluering av vurderingen eller vurderingsordningene.<sup>9</sup>

---

<sup>6</sup> *Thematic review of the first years of tertiary education*. Country note: Norway. (OECD Publications).

<sup>7</sup> Ibid.:17.

<sup>8</sup> *Assessment Principles*, University of Queensland, Australia.

<sup>9</sup> *Norsk Lovtidend*, I. nr 6(2002).

“Nye evalueringsformer” blir på engelsk gjerne omtalt som “innovative assessment”, eller “authentic assessment. G. Mowl beskriver dette på følgende vis: “Innovative assessment is about empowering learners. Learning rather than assessment should be the end product of higher education. ... Students should learn through assessment, not learn to be assessed.”<sup>10</sup> I en meget kjent studie fant amerikaneren Benson Snyder at det ved siden av den offisielle læreplanen eksisterer en skjult læreplan, med konsekvenser for studieatferd og læring:

As I examined colleges, universities, and secondary schools, I was struck repeatedly with the importance of a hidden agenda, a hidden curriculum, which determined to a great degree the way in which the various participants played the game, read the cues, adapted to their immediate educational circumstances.<sup>11</sup>

Vurderingsordningene definerer langt på vei den skjulte læreplanen. Rowntree uttrykker dette på følgende måte: “If we wish to discover the truth about an educational system, we must look into its assessment procedures.”<sup>12</sup> Tidligere eksamensoppgaver forteller indirekte hva som blir oppfattet som viktig og mindre viktig: “Jo mindre faget er, desto sikrere kan en være på at løsning av gamle eksamensoppgaver gir god uttelling.”<sup>13</sup> Det er altså ikke alltid pensum som former eksamen, men eksamen som langt på vei former pensum:

Alle eksamener prøver sikkert å teste elevenes forståelse, men vanligvis er det måter å lure seg unna på. Løse gamle eksamenssett, eller lære seg litt metoder og sånn, uten egentlig å ha forstått særlig mye.<sup>14</sup>

Det gis altså gode grunner for å planlegge eksamens- og vurderingsformene som sentral del av et undervisnings- og læringsopplegg. Faglærere i høyere utdanning har vært mer opptatt av å planlegge undervisning som en aktivitet i og for seg, mens eksamen og vurdering var noe som kom i tillegg ved slutten av semesteret. I noen grad jobber imidlertid studenter baklengs gjennom pensum ved først å undersøke når og hvordan de blir vurdert, slik oversikten

---

<sup>10</sup> Mowl (s.a.).

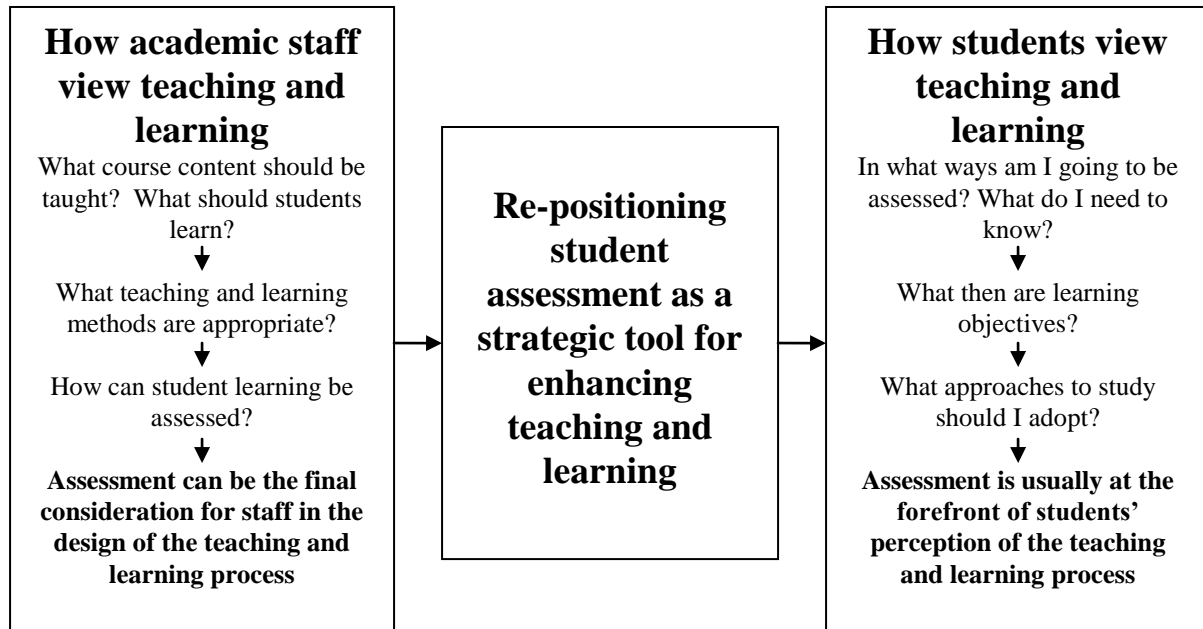
<sup>11</sup> Snyder (1971:xii).

<sup>12</sup> Rowntree (1987:1).

<sup>13</sup> Student NN, NTNU, 1997.

<sup>14</sup> Student NN, NTNU, 1997.

viser.<sup>15</sup>



Hva mener vi så med “nye” vurderingsformer, og hvordan kan vi skape evalueringsopplegg som bidrar til å fremme ønsket læring? Å lykkes i dette handler ikke bare om en instrumentell bruk av motepregete former. Slik jeg ser det, handler det om å komme på sporet av en ny tenkemåte, slik Harris og Bell beskriver: “It is not the actual methods or tools of assessing which we believe should be changed in many cases, rather the underlying philosophy and the aims of their use and application.”<sup>16</sup> En samlet oversikt over nye vurderingsformer finnes ikke, og det er sammenstillingen og bruken av ulike former som oftest vil være interessant.

- Prosjekter
- Gruppeoppgaver
- Hjemmeeksamen
- Flervalgseksamen
- Åpen bok-eksamen
- Muntlige presentasjoner
- Porteføljer eller “mapper”
- Simuleringsøvelser og rollespill
- Egen-, kollega- og medvurdering
- Realistiske eller reelle (real life) oppgaver

<sup>15</sup> James & Devlin (2002:8).

<sup>16</sup> Harris & Bell (1990:97).

- Ulike typer skriveoppgaver for spesielle målgrupper og formål

Kvalitetsreformen nevner studentporteføljer/mappevurdering og deleksamener som spesielt hensiktsmessig som kombinasjon mellom undervisning og evaluering. Departementet er opptatt av at studentene skal få hyppigere tilbakemeldinger på egne arbeid, og at de skal jobbe jevnere gjennom semesteret, slik at kortsiktig pugg foran slutteksamen blir unngått.<sup>17</sup> Det som forener det vi her har omtalt som “nye” eksamensformer, er først og fremst ønsket om å forbedre kvaliteten i studentenes læring. I praksis betyr det at evalueringen inngår som en integrert del av det løpende undervisnings- og læringsarbeidet. Eksamen og evaluering blir et virkemiddel for å lære, ikke bare et middel for å kontrollere læringsresultatene. For studentene blir dermed evaluering også et redskap for egen læring. Bruk av læringsorienterte evalueringsformer betyr derfor en maktforskyvning fra faglærer og sensor til studentene. Nye eksamensformer kan dermed bidra til at studentene i større grad tar ansvar for sin egen læring.

Men hvordan kan nye eksamensformer bidra til bedre læring? Forskning viser at det i noen studier er mulig å få en god karakter uten jevn innsats, men med et stormløp foran eksamen, slik denne studenten i sivilingeniørstudiet beskriver: “Det er ofte man får god uttelling ved å ha regnet mange eksamensoppgaver, og nesten ikke åpnet pensumboken.” Den langsiktige læringseffekten av slik studieatferd er imidlertid begrenset. Ved å knytte evaluering og læring sammen åpnes muligheter for mer sammenhengende læringsarbeid. For eksempel er det mulig å fremme bestemte typer kompetanse ved å aktivisere studentene i ulike typer arbeid gjennom semesteret. Slikt arbeid kan organiseres i ulike former, alene eller i grupper, der prosessen og resultatet kan tjene flere formål samtidig, både for læring og som bidrag til sertifisering.

Det største potensialet for bedre læring ligger trolig i en annerledes og mer effektiv bruk av studentene selv. Å vurdere egen og andres besvarelser kan være en utmerket læremåte, fordi studentene må kunne begrunne og utdype sitt eget syn overfor medstudenter og eventuelt faglærer. Dette likner på arbeidsformen i en akademisk disputas, der målet er søken etter sann kunnskap mer enn det er en kamp om å ha eller få “rett”. Under en hjemmeeksamen i sivilingeniørstudiet der studentene kunne jobbe sammen i par, skrev en av dem følgende:

---

<sup>17</sup> *St. meld nr. 27* (2000-2001: kapittel 5.3.5)



Denne læringsmetoden har klart noe for seg. Læringseffekten er høy. Lærte spesielt mye av å diskutere med andre, ta et standpunkt som må forsvares. Man trenger lengre ned i stoffet på den måten. Håper dette er noe som er kommet for å bli.<sup>18</sup>

Tradisjonell eksamen setter grenser for hvilke typer kompetanse som lar seg kontrollere, og dermed for hvilke læringsaktiviteter som blir tillagt vekt i læringsforløpet. Studenter opplever at det ikke “lønner” seg å lære stoff som ikke blir opplevd som relevant for eksamen. Sentrale mål for nye eksamensformer kan skisseres i noen hovedpunkter:

- Å fremme motivasjon og engasjement for egen læring
- Å fremme en sammensatt, relevant og helhetlig kompetanse
- Å utvikle en selvkritisk holdning og evne til egenvurdering
- Å gi studentene et bredere grunnlag for å utvikle egen kompetanse
- Å bidra til at *studentene lærer seg*, ikke bare mottar kunnskap fra andre

Hensikten med nye eksamensformer er å skape et bedre læringsmiljø, og mer meningsfulle læringsopplevelser. Dette forutsetter at studentene blir mer aktive og ansvarlige for sin egen læring. Det som ved første øyekast ser ut som en enkel teknisk endring, krever en ny tilnærming til egen rolle enn det de fleste faglærere og studenter er kjent med fra tradisjonell eksamen.

#### **4. Ulike tilnærminger til læring**

Gis det så noen sammenhenger mellom læringsprosesser og læringsresultater? På 1970-tallet undersøkte Marton og Säljö dette spørsmålet eksperimentelt blant universitetsstudenter i Sverige. Her ble det dokumentert at studentene rettet oppmerksomhet mot ulike aspekter ved lesning av en felles tekst. Studentene hadde ulik oppfatning av hva utfordringen egentlig besto i og satt etter gjennomlesningen igjen med ulik oppfatning av hva teksten egentlig handlet om.

It was found that in each study a number of categories ... containing basically different conceptions of the content of the learning task could be identified. ... If there are qualitative differences in learning it seems very

---

<sup>18</sup> Student NN, NTNU, 1997.

likely that there are corresponding differences in the process of learning, i.e., in the way people set about learning.<sup>19</sup>

In the case of surface-level processing the student directs his attention towards learning the text itself (the sign), i.e., he has a 'reproductive' conception of learning which means that he is more or less forced to keep to a rote-learning strategy. In the case of deep-level processing, on the other hand, the student is directed towards the intentional content of the learning material (what is signified), i.e., he is directed towards comprehending what the author wants to say about, for instance, a certain scientific problem or principle.<sup>20</sup>

Gibbs legger vekt på at læringsprosesser først og fremst er styrt av studentenes intensjoner, og at læringsresultatene oftest er et produkt av dette og sjeldnere skyldes intellektuell evne.

Every lecturer has experienced students who seemed determined to give back, in essays and reports, exactly what they were given, in lectures. Other students strive to develop their own perspectives and synthesis of the subject. This may sometimes involve a difference in ability, but most often it involves a difference in intention: students are trying to achieve different things. These two extreme intentions have been termed a surface approach and a deep approach.<sup>21</sup>

Sammenhengen mellom læringsstrategi og resultat er beskrevet av Biggs på følgende vis:

Relationships between such approaches and outcome are exceptionally strong: studies ... show that a surface approach, almost without exception, leads to a quantitative outcome of unstructured detail, and a deep approach to an appropriately structured outcome.<sup>22</sup>

Martons og Säljös studier ble utført under betingelser som ikke helt kan sammenliknes med reelle undervisnings- og læringssituasjoner der vissheten om *eksamen* også influerer på studieatferd, slik en student skriver "... jeg håpet i bunn og grunn på en lettkjøpt karakter." Studenter med en utpreget strategisk tilnærming til sin læring legger vekt på det som under gitte betingelser "lønner

---

<sup>19</sup> Marton & Säljö (1976).

<sup>20</sup> Ibid.

<sup>21</sup> Gibbs (1992: 2).

<sup>22</sup> Biggs (s.a.) (Her sitert fra Gibbs, G., op. cit.)

seg” for å oppnå en best mulig karakter med minst mulig innsats. I enkelte tilfelle framstår læringsprosessen som en parodi av det den var ment å være. Martons og Säljös opprinnelige kategorier er derfor blitt supplert med en tredje tilnærming, som er interessant ved institusjonsbasert læring:

... other researchers ... have identified an achieving or strategic approach in which students adopt an expedient position and use either a surface or deep approach depending on pressures of time or assessment expectations.<sup>23</sup>

Alle studenter har prinsipielt sett sin egen måte å studere på. Vel vitende om dette, ønsket jeg under mitt doktorgradsarbeid i sivilingeniørstudiet ved NTNU å undersøke om det var mulig å identifisere studieatferd som beskrevet i de tre kategoriene jeg har omtalt foran. Individuell studieatferd sees her som resultat av en intern tolknings- og forhandlingsprosess på grunnlag av “ytre rammer” (som forelesninger, øvinger og eksamen) og egen praksisteori, det vil si de “brillene” studentene bruker for meningskonstruksjon og handling. Under samtale med et lite utvalg studenter (11 stk.) ba jeg hver enkelt beskrive egne intensjoner underveis i studiet.

#### **Eksempel på overflatisk læring (Surface Approach):**

Prøvde liksom å lære alt litt, og ingenting grundig. ... Jeg føler at jeg ikke “har taket” på noen ting. ... Brydde meg ikke så mye om å lære ting grundig, men satset på en mer overflatisk forståelse. Følte egentlig at jeg manglet total kontroll.<sup>24</sup>

#### **Eksempel på dyplæring (Deep Approach):**

Jeg er innstilt på å lære noe ... Dette skal være kunnskap som sitter. Jeg går ikke her bare for å gjøre det bra til eksamen. Derfor har jeg kanskje ikke fått så bra på selve eksamen som mange andre som bare har pugget eksamensoppgaver har gjort.<sup>25</sup>

#### **Eksempel på taktisk/strategisk læring (Achieving/Strategic Approach):**

Motivasjonen min var lav gjennom hele semesteret ... Øvingene har i høyeste grad vært preget av bevisstløs plikt. Jeg har overhodet ikke fått noen

---

<sup>23</sup> Boud (1995:26).

<sup>24</sup> Intervju med student NN, NTNU, 1997.

<sup>25</sup> Intervju med student NN, NTNU, 1997.

økt forståelse ... Eksamensforberedelsene var utelukkende preget av at jeg ville ha en best mulig karakter, uten tanke på egen læring.<sup>26</sup>

Beskrivelsen av de tre kategoriene representerer her ikke egenskaper ved studentene, men konkret atferd framtrer i møte mellom student og kontekst. Det er ikke gitt at flertallet av studentene uten videre faller inn i noen av de kategoriene vi har omtalt, men oftest vil en kategori være mer beskrivende enn en annen. Dersom vi ønsker å forbedre læring, trenger vi å kartlegge sentrale aspekter ved konteksten, og ved studentenes opplevelse av den. Ved å forandre de ytre “rammene” for læring, kan vi sannsynliggjøre andre prioriteringer fra studentenes side. Derfor er det så viktig at vi kan kartlegge hvilke forhold som har en spesielt handlingsdrivende karakter for studentene. Studentevaluering av undervisning, som ensidig retter fokus mot faglærers presentasjon, vil neppe påvirke valg av studietilnærming i særlig grad. Vurderingsformene har, sammen med flere forhold, i så måte et langt større potensial.

## 5. En handlingsteoretisk tilnærming

Spørsmål knyttet til atferd og atferdsendring har vært gjenstand for en rekke studier. Berglind hevder at samfunns- og humanvitenskapene har vært dominert av to hovedtyper av forklaringsmodeller, nemlig de *kausale* og de *teleologiske*.<sup>27</sup> Iblant brukes begrepene *finale* og *intensjonale* forklaringstyper. Begrepet *teleologi* kommer fra gresk, og betyr læren om det formålstjenlige.<sup>28</sup> Dette er et filosofisk standpunkt som hevder at hendelser og virkende krefter i verden ikke bestemmes av rent mekanistiske årsaker, men av hensiktmessighet eller formålstjenlighet. Von Wright relaterer de to modellene til henholdsvis Galilei og Aristoteles.<sup>29</sup> Kausale forklaringer, eller årsaksforklaringer, er ifølge Von Wright et karakteristisk trekk ved positivismen. Individuelle tilfelle sees her som resultat av mer generelle lover. Handlinger som forsøkes forklart ut fra aktørens hensikt eller mål, har enten vært sett på som uvitenskapelige, eller forsøkt tvunget inn i mønstre for kausale forklaringer.

Sosiologen Boudon ser menneskets handlinger i lys av et sosialt interaksjonssystem. Han viser både til de ytre rammene for menneskelig handling og til menneskenes valg innen disse rammene.<sup>30</sup> Den som i nordisk sammenheng har bidratt mest til den “ramfaktorteoretiske” tenkningen er

---

<sup>26</sup> Intervju med student NN, NTNU, 1997.

<sup>27</sup> Berglind (1995:13)

<sup>28</sup> *Gyldendals fremmedordbok* (1965:268).

<sup>29</sup> Von Wright (1993:2).

<sup>30</sup> Boudon (1981:16).

professor Urban Dahllöf ved Uppsala Universitet. Dahllöfs modell forutsetter en sammenheng mellom strukturelle vilkår og strukturering av handling.<sup>31</sup> En konsekvens av dette var at undervisning og læring kunne analyseres både ut fra et aktørperspektiv og et strukturelt perspektiv. Dahllöfs vesentligste bidrag består, etter min mening, i en modell for analyse og planlegging av undervisnings- og læringsforløp. Den kan skisseres på følgende vis:



Dette er en prinsipielt åpen modell. Avhengig av problemstilling vil spørsmål om rammer, prosesser og resultater, kunne stilles på høyst ulike vis. Som Dahllöf selv har gjort rede for, er det meningsløst å stille et generelt spørsmål om hva som er rammer, prosesser og resultat.<sup>32</sup>

Om utgangspunktet är någon form av ram, är det lämpligt att först tänka efter i hvilket syfte den har tillkommit och då fixera en dimension för avsett utfall för att därefter fråga vilka slag processer som det i första hand blir aktuellt att analysera. Detta utesluter inte att också sidoeffekter och oavsiktliga processer studeras, men dessa måste ändå komma i andra hand.<sup>33</sup>

På en rekke felt, ikke minst i profesjonsutdanninger, er det sterkt ønskelig å utvikle gode læringsprosesser for å sikre bestemte typer kompetanse. Her vil vi kunne bruke modellen til analyse av forholdet mellom rammer og prosesser. Vi kan nyansere og utvikle rammebegrepet slik at det gir mening i konteksten. Dersom det ikke lar seg gjøre, er naturligvis ikke modellen relevant for problemet. Hvis vi retter oppmerksomhet mot resultater i form av kunnskaper, ferdigheter, holdninger, gjennomstrømning og lignende, er det ikke alltid lett å spore årsakene til resultatet. "I så fall behöver båda processlänkar och eventuellt styrande ramar fixeras och en teori på den grunden utvecklas som förklarar ett mindre gott resultat. ..."<sup>34</sup> Dette er en sentral utfordring ved teoribasert evaluering av undervisning og læring, der vi ønsker en dypere forståelse av *sammenhenger* mellom rammer, prosesser og resultater. Dahllöf legger vekt på forskjellene mellom en modell og en teori. En pedagogisk teori forutsetter

... en detaljerad och uttömmande beskrivning av mekanismerna i en process, som på ett konsistent och övertygande sätt ger förklaringsvärde åt

---

<sup>31</sup> Dahllöf, U. (1997). Trykt versjon i Kvalsund, R. m. fl. (1998)

<sup>32</sup> Ibid.: 8.

<sup>33</sup> Ibid.: 7.

<sup>34</sup> Ibid.: 7.

relationerna mellan de aktuella ramförutsättningarna, förloppen i undervisningen och i den inlärning som i sin tur mynnar i resultat på kunnskapsprov i olika kursmoment. Det för teorin specifika är altså den detaljerade analysen av mekanismerna i de aktuella processerna.

En modell kan imidlertid være svært nyttig også i de tilfelle der vi ikke makter å utvikle en teori i tråd med Dahllöfs definisjon. Teoretiske ansatser og partielle teorier for analyse av bestemte fenomen eller sammenhenger er oftest også nyttig. Under utviklingsarbeid handler det om å komme på sporet av sammenhenger som grunnlag for forbedringer. Svært mye evaluering av undervisning og læring i høyere utdanning i Norge har vært gjennomført uten mer omfattende teoretisk forankring, eller mer selvstendige forsøk på teoribygging.

## 6. Rammer i praksis

Noen vil oppleve at de begrepene jeg har omtalt foran fortsatt er diffuse. Under et utviklingsarbeid har vi interesse å vite hvilke trekk ved undervisning og studieopplegg som hemmer eller fremmer bestemte typer studieatferd og læring. Slik innsikt kan i sin tur fungere som beslutningsunderlag for hvilke tiltak som bør settes ut i livet. Som eksempel konkretiserer Gibbs følgende trekk som fremmer overflatisk læring (Surface Approach):<sup>35</sup>

- A heavy workload
- Relatively high contact hours
- An excessive amount of course material
- A lack of opportunity to pursue subjects in depth
- A lack of choice over subjects and a lack of choice over method of study
- A threatening and anxiety provoking assessment system

Rammebegrepet kan også eksemplifiseres ved et par sitater fra sivilingeniørstudenter ved NTNU. Under et intervju med undertegnede falt følgende kommentarer fra to av studentene:

Jeg synes vi skulle hatt litt flere oppgaver som gikk på teori. Da måtte du sette deg litt mer inn i det, og hvis du tar teorioppgaven først, så vet du i hvert fall hva du regner på. Jeg synes det er veldig mye regning, regning, regning hele tiden. Det ble til at jeg regnet uten å forstå. ...<sup>36</sup>

---

<sup>35</sup> Gibbs (1992:9).

<sup>36</sup> Intervju med student NN, NTNU, 1998.

De skulle gått ned i vanskelighetsgraden på oppgavene og hatt flere oppgaver, slik at vi hadde fått repetert basisprinsippene gang på gang. Du føler at du fikk til noe, men du behersket det overhodet ikke. Du har akkurat med et hår klart det med litt hjelp. Du føler at du er utilstrekkelig. Du har ikke fått innlært begrepene godt nok.<sup>37</sup>

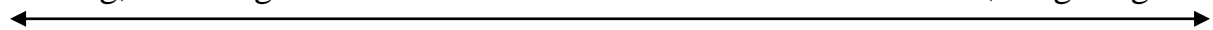
Som vi forstår, er rammebegrepet tøyelig, men i motsetning til undervisning i tradisjonell forstand har vurderingsopplegget blitt sett på som en sak for seg. Den bærende ide bak ønsket om mer varierte og prosessrettete evalueringsformer er ønsket om dypere og mer relevant læring. Studentene har imidlertid fri vilje til å gjøre sine egne valg. Vi kan aldri manipulere studenter i negativ betydning, men derimot styrke og utvikle rammer som erfaringsmessig sett øker sjansen for at ønsket om læring og kompetanseutvikling finner sted. Dette er en dynamisk prosess der alle parter på ulike vis er ansvarlig for å bidra, men det forutsetter ører som hører og øyne som ser, altså evne og vilje til dialog mellom faglærer, studenter og administrasjon.

## 7. Summativ eller formativ evaluering?

Under pedagogisk utviklingsarbeid vil vi ha stor nytte av noen få sentrale begreper. Det gjør det lettere å delta i en konstruktiv dialog med kolleger og andre om det vi holder på med. La meg først introdusere begrepene *formativ* og *summativ* evaluering. Fra norsk høyere utdanning kjenner vi slutteksamen først og fremst som et kontrolltiltak, altså som summativ evaluering. I blant ønsker vi imidlertid også å benytte vurderingsordningene som del av et læringsopplegg, oftest omtalt som formativ bruk. Porteføljebasert vurdering er et eksempel på dette. I noen tilfelle står valget ikke mellom den ene eller den andre formen, men om en kombinasjon. Derfor er det naturlig å skille mellom vurderingsformer og vurderingsopplegg. Figuren nedenfor viser hvordan et evalueringsopplegg for et kurs kan plasseres langs en linje med ekstremalpunkter, som markerer *evaluering for læring* og *evaluering for kontroll*.

Læring, utvikling

Kontroll, rangering



### Ulike dimensjoner ved evaluering

Vurderingsordningene i høyere utdanning har, i ly av lov- og regelverk, langt på vei levd sitt eget liv. De har vært en del av den tause kunnskapen blant fag-

---

<sup>37</sup> Intervju med student NN, NTNU, 1998.

lærerne, og har dermed kun i sjeldnere tilfelle vært gjenstand for refleksjon og aktive valg. Dette har i høy grad bidratt til at krav om begrunnelser og systematiske studier av effekter av vurderingsopplegg har uteblitt. Aktiv bruk av evaluering som del av et læringsforløp forutsetter en mer reflektert holdning fra faglærernes side med sikte på å kunne utvikle gode evalueringsdesign. Dette setter i sin tur krav om begrunnelser overfor en selv, kolleger, administrasjon og ikke minst studentene. Spørsmål om valg av evalueringsformer, kan med hell diskuteres med studentene. Det handler om å kunne sette hensiktsmessige rammer for læring, og om kriterier for sertifisering. Dette dreier seg ikke bare om valg av vurderingsformer, men om oppgavetyper, bruk av litteratur og annet. At studenter har et engasjement i spørsmål om vurderingsformer er tydelig, slik følgende sitat viser:

Jeg oppfatter at eksamen nesten utelukkende som et slags mål eller rangering av studentene. Jeg vil nesten kalle det en slags intellektuell utvelgelse. Jeg synes ikke eksamen stimulerer til best mulig læring utover det å oppnå best mulig karakter med sikte på seinere jobbmuligheter.<sup>38</sup>

I min undersøkelse fra sivilingeniørstudiet, er det likevel interessant å se at det finnes mer nyanserte syn. Derfor blir det feil å hevde at en gitt eksamensform “mekanisk” bidrar til én bestemt studieatferd. Studentene tolker signaler i det de opplever som sitt “handlingsrom”, prioriterer og skaper selv sin individuelle studieatferd, slik denne studenten beskriver:

Eksamen er med på å disiplinere mine studievaner. Det er klart karakterer brukes til å rangere studenter for framtidige arbeidsgivere, men i alt målrettet arbeid en driver trenger man en tidsfrist og et definert mål for hva man skal oppnå. Derfor mener jeg eksamen også stimulerer til læring.<sup>39</sup>

I min undersøkelse blant studenter i sivilingeniørstudiet hevdet et flertall av respondentene at *karakterene* hadde en spesielt sterkt motiverende effekt. Kun et lite antall trodde de ville lært like mye og like godt dersom det bare ble gitt *bestått* og *ikke bestått* til eksamen. Sivilingeniørstudiet benyttet på det aktuelle tidspunktet en karakterskala fra 1,0 til 6,0, med 4,0 som dårligste ståkarakter. Undersøkelsen min tyder på at det først og fremst er de graderte karakterene som virker motiverende for studentenes innsats, slik sitatene nedenfor beskriver:

---

<sup>38</sup> Student NN, NTNU, 1997.

<sup>39</sup> Student NN, NTNU, 1997.



Hvis det kun hadde vært om å gjøre å bestå, hadde jeg aldri i verden gidde å lese så mye. Jeg ville heller lest på andre fag. Det handler om **poeng**.<sup>40</sup>

Karakterer er motiverende i seg selv, men jeg tror ikke den dype forståelsen kommer på grunn av dette.<sup>41</sup>

Da amerikaneren Benson Snyder introduserte begrepet *den skjulte læreplanen*, bidro han til et nytt perspektiv for studier av høyere utdanning. En konklusjon fra Snyders undersøkelser var at studentenes prioriteringer og valg var lite sammenfallende med intensjonene i den *offisielle* læreplanen.<sup>42</sup> Innen nyere forskning blir eksamensformene trukket fram som spesielt viktig som grunnlag for studieatferd og læring. Eksamensordningene forteller om de verdier og den pedagogiske tenkning som ligger til grunn for en utdanning. I hvilken grad blir for eksempel evaluering sett på som en integrert del av et læringsforløp? Eventuelt i hvilken grad er dette noe som kommer i tillegg til, og faller utenom et ordinært undervisnings- og læringsforløp?

## 8. Pålitelighet og gyldighet

Eksamen og sensur er viktig for studentene. Å bestå eksamen er et mål i seg selv, og for mange studenter er det knyttet et konkurranseelement ved den interne rangeringen mellom kandidatene. Derfor er spørsmål knyttet til *pålitelighet* et viktig tema ved vurdering<sup>43</sup>, mens vurderingens *gyldighet* har vært et mindre framtrædende emne. Vi sier gjerne at vurderingen er pålitelig hvis resultatet ikke i vesentlig grad avhenger av hvem den er utført av. Det er spesielt to former for pålitelighet vi er opptatt av. På den ene siden at vurderingens resultat ikke blir vesentlig forskjellig dersom den utføres av to uavhengige bedømmere, og på den andre siden at resultatet ikke blir vesentlig forskjellig dersom det utføres av én og samme person på to ulike tidspunkt. Pålitelig vurdering forutsetter kriterier for hvilke aspekter ved en oppgave som skal tillegges vekt, og hvordan kriteriene skal anvendes i hvert enkelt tilfelle.

Ett tema som ofte går igjen er spørsmål knyttet til *intersubjektiv* enighet mellom to ulike bedømmere. Gir for eksempel faglærer samme vurdering av besvarelsene, eller er det store sprik fra kandidat til kandidat? I min undersøkelse av eksamen og sensur i ett grunnkurs i sivilingeniørstudiet var det

---

<sup>40</sup> Student NN, NTNU, 1997.

<sup>41</sup> Student NN, NTNU, 1997.

<sup>42</sup> Snyder (1971).

<sup>43</sup> Kvale (1970)

*signifikant korrelasjon* mellom faglærer og sensor i de fire årene som ble undersøkt. Det betyr ikke nødvendigvis at sensor og faglærer har gitt samme karakter, men at det er en svært høy grad av *systematisk samvariasjon* mellom sensors og faglærers vurdering. Middelveien av sensors vurdering for kandidatene var dårligere enn faglærers vurdering. Årsaken kan være at vurderingskriteriene ikke har vært tydelige nok, eller at faglærer og sensor har benyttet kriteriene på ulik måte. Fordi de endelige karakterene i faget ble fastsatt etter et forhåndsdefinert nivå, var middelveien av differansen mellom sensors og faglærers vurdering i seg selv ikke vesentlig. Det avgjørende var forskjellen ved vurdering fra kandidat til kandidat. Nær 85 % av kandidatene ville fått samme karakter uten sensors bidrag.

Innen psykologisk testing har det vært lagt vekt på å utvikle metoderegler for å fremme en pålitelig vurdering. Den første regelen gjelder hvilke aspekter ved prestasjonen det skal tas hensyn til. Den andre regelen gjelder spørsmålet om hvor mye ulike aspekt ved besvarelsene skal telle ved den helhetlige vurderingen. Den tredje regelen etterspør standarder for hvordan en karakter-skala skal benyttes. Den fjerde metoderegelen etterspør systematisk opplæring og trening for dem som står ansvarlig for vurderingen. I vårt land har vi ikke hatt tradisjon for noen form for obligatorisk kvalifisering for vurdering og sensur. Kompetanse på dette feltet har grad vært ervervet i praktisk sensurarbeid slike det har vært tradisjon for i ulike miljø.

Pålitelig vurdering er et viktig mål i seg selv, men kan i sin ytterste konsekvens føre til at bare det som kan telles eller måles blir interessant. I dagens situasjon, med et sterkt ønske om å vektlegge læringsaspektet ved vurdering, er pålitelighet som eneste fokus utilstrekkelig. Vi innser at kunnskapens former er mangfoldige, at den varierer fra fag til fag, og at all vurdering er basert på skjønn og dermed beheftet med usikkerhet. Vi er derfor også opptatt av at vurderingen skal være *gyldig*. Dette er imidlertid ikke et entydig begrep, og vi utdyper her tre sentrale aspekter, nemlig *innholdsgyldighet*, *prediktiv gyldighet* og *konsekvensgyldighet*.

Hvordan kan vi se for oss forholdet mellom innholdet i et fag og vurderingsopplegget i faget? Eller sagt på en annen måte: Hvordan er sammenhengen mellom læringsmålene og eksamen? Vurderingen er gyldig med hensyn til innhold dersom den dekker flest mulig av de sentrale læringsmålene i faget. Dersom vurderingen skjer på basis av få og/eller perifere læringsmål, er den i liten grad gyldig med hensyn til innhold. Slutteksamen som form har klare begrensninger med hensyn til hvilke aspekter ved et fag som lar seg kontrollere eller måle. Derfor ser en ikke sjelden at bare enkelte av læringsmålene blir

prøvet ved eksamen. Uten systematiske undersøkelser tar vi neppe mye feil ved å hevde at det er pålitelighetsaspektet ved vurdering som er blitt tillagt størst vekt i norsk høyere utdanning. Vi vil derfor i mange tilfelle finne vurderingsopplegg med stor vekt på pålitelighet, men med relativt liten vekt på gyldighet.

Spesielt i profesjonsstudier er *prediktiv gyldighet* ved vurderingsordningene et mål. Resultatet av en summativ vurdering forutsettes å fortelle noe om kandidatene med relevans for et yrke. Det betyr at vurderingsoppleggets form og innhold bør speile flest mulig av de krav til kompetanse som blir stilt til en profesjonell yrkesutøver innen fagområdet. Det kan for eksempel være en legestudent som skal demonstrere kommunikasjonsferdigheter med en pasient, eller en sykepleiestudent som viser ferdigheter i å sette en sprøyte. Spørsmål knyttet til prediktiv gyldighet ved allmennvitenskapelige fag, som ikke sikter mot et spesielt yrke, er mer komplisert. Men også her er det mulig å lage vurderingsopplegg som prøver kandidatenes evne til refleksjon og forståelse av et fagstoff, i motsetning til ren memorering av pensum.

Begrepet *konsekvensgyldighet* er trolig lite kjent for de fleste, men like fullt et svært viktig aspekt ved vurdering. La meg bruke en tradisjonell slutteksamen som eksempel. Dens viktigste funksjon er summativ, den brukes oftest for kunnskapskontroll og sertifisering av kandidater, for eksempel innen et konkret fag eller kurs. På samme tid har en slik eksamen formative konsekvenser ved at den påvirker når, hva og hvordan studentene lærer. La meg illustrere dette med et fag fra sivilingeniørstudiet ved NTNU. Et marinteknisk grunnkurs i tredje studieår hadde ca. 100 studenter. Undervisningen var basert på forelesninger og øvinger, og stipulert tidsforbruk per uke i faget var 12 timer. En enkelt tidsstudie på grunnlag av et lite utvalg studenter viste imidlertid at det gjennomsnittlige tidsforbruket var mindre enn stipulert gjennom hele semesteret, bortsett fra to uker før eksamen. Skippertakspreget innsats var svært utbredt. Eksamen hadde altså en åpenbart tilbakevirkende kraft på studieatferd og læring i faget.

At et vurderingsopplegg er gyldig i sine konsekvenser, handler både om når og hvordan studentene arbeider. Det finnes typer av læring og kompetanseutvikling som krever innsats over tid for å etablere nødvendig faglig modning. Å ta stilling til valg av vurderingsformer ut fra en slik tilnærming forutsetter at fagansvarlig for seg selv har en tydelighet om arten av kompetanse som ønskes utviklet. I tillegg forutsettes innsikt i hvilke arbeidsformer som kan være hensiktsmessig for å fremme bestemte typer kompetanse. Dette representerer en handlingsteoretisk tilnærming, der “rammene” gir incentiver til bestemte typer studieatferd. I flere år var vurderingsformene et ikke-tema som virkemiddel for

læring. I stedet ble det lagt vekt på andre tiltak, som kurs i studieteknikk, lesetrening ol. Dette var imidlertid langt på vei som olje på vannet i forhold til strukturer som påvirker studieatferd og læring i ethvert studium. Jeg ønsker med dette ikke å si noe galt om denne typen kurs i og for seg. Poenget er at slike tiltak neppe i seg selv har en handlingsgenererende kraft som motstår andre sterke impulser.

## 9. Normreferanse og kriteriereferanse

De to begrepene refererer til vidt forskjellige prinsipper for vurdering og karaktersetting. Ved en normbasert vurdering er ikke den faglige prestasjonen i og for seg avgjørende, men de retningslinjer karakterene reguleres i forhold til. La meg illustrere dette med et eksempel. Ved første gangs vurdering i et fag var middelverdien av sensors og faglærers vurdering 3,5 på en skala fra 1,0 til 6,0, med 4,0 som dårligste ståkarakter. Det var imidlertid tradisjon for at middelverdien av karakterene i faget lå nær 2,5. Faglærer og sensor mente det hadde noe med rettferdighet å gjøre at det ikke ble gjort store endringer på dette fra år til år. Resultatet var i dette tilfellet at samtlige karakterer ble forbedret med en hel karakter. Eksemplet viser hvor nødvendig det er å kjenne prinsippene for karaktersetting for å kunne vurdere hva en karakter forteller, og for å kunne sammenlikne karakterer fra fag til fag. Normbasert vurdering kan, som i eksemplet, være knyttet til en gitt middelverdi for karakterene, men prinsippene kan være langt mer sofistikerte. For bruk ved et stort antall kandidater blir det gjerne laget modeller for prosentvis fordeling av karakterene, for eksempel i form av en normalfordeling.

*Kriteriereferanse* setter, i motsetning til normreferanse, konkrete krav til prestasjoner. Hva studenten får i karakter, beror på egen prestasjon. Innen en rekke yrkesutdanninger blir det stilt absolutte krav til ferdigheter som det ikke gis grunn for å slakke på. Ett aktuelt eksempel er tannlegeutdanningen der det blir stilt relativt strenge krav til kompetanse før sertifisering til yrke. Vurdering på grunnlag av kriteriereferanse vil langt på vei “styre” pensum, og åpner for en beskrivelse av hva som er oppnådd av kompetanse i studiet. Kriteriebasert vurdering vil prinsipielt kunne fungere godt sammen med formativ vurdering, fordi den bygger opp om en agenda for diskusjon om kompetanse. Undervisningsformer og eksamen bør derfor være *gyldige*, det vil si at de bør bidra til å fremme ønsket kompetanse i kurset eller studiet.

Ved Norges teknisk-naturvitenskapelige universitet eksisterte det inntil nylig ingen uttalte retningslinjer for karaktersetting. I det faget jeg undersøkte i mitt

doktorgradsarbeid, viste det seg at karakterene først ble fastsatt på grunnlag av en kriteriereferanse, altså i forhold til bestemte faglige krav. Etter dette ble karakterene justert i forhold til en ønsket middelvei, altså en normbasert vurdering. Sensor i det aktuelle faget beskriver dette på følgende vis:

Det man kanskje gjør, er å justere karakterene litt ned slik at man får et fornuftig snitt i forhold til det som har vært gitt før. At man føler det som urettferdig at studentene i år skal få dårligere snitt enn studentene i fjor.<sup>44</sup>

Hvilke aspekter ved besvarelsene som skal telle i vurderingen, er heller ikke gitt: "Det er noe som går seg til etter hvert."<sup>45</sup> Standarder for hvordan en tallskala skal benyttes utvikles gjennom praksis og bidrar til ulike vurderingskulturer innen ulike fag i samme studium.

## 10. Lovverket for høyere utdanning

Lov om universiteter og høyskoler av 12. mai 1995, med senere endringer, setter rammer for hvordan eksamen og vurdering skal foregå, og loven presiserer studentenes rettigheter med hensyn til klage på sensurvedtak. Paragraf 50 nr. 1 og 2 lyder nå som følger:<sup>46</sup>

1. Styret skal sørge for at studentenes kunnskaper og ferdigheter blir prøvet og vurdert på en upartisk og faglig betryggende måte. Vurderingen skal sikre det faglige nivå ved vedkommende studium. Det skal være ekstern vurdering av vurderingen eller av vurderingsordningene.

2. Styret oppnevner sensor ved eksamen, prøve, bedømmelse av oppgave eller annen vurdering når resultatet inngår på vitnemålet eller innregnes i karakter for vedkommende studium. Det skal være minst to sensorer, hvorav minst en ekstern, ved bedømmelse av kandidatens selvstendige arbeid ved høgre grad.

Sett i forhold til tidligere versjon av loven, er endringene ikke spesielt store. Formuleringene gir assosiasjoner til det kontrollparadigmet som lå til grunn for blant annet en svært utbredt bruk av individuell slutteksamen. At vurderingsordningene kan tjene andre funksjoner enn kontroll og sertifisering, har ikke kommet med verken i lovteksten eller i merknadene. Det prinsipielt nye er at institusjonen selv bestemmer hvordan ekstern deltakelse i vurderingen skal finne sted. Det kan være ekstern deltakelse ved den enkelte vurdering, eller

---

<sup>44</sup> Personlig meddelelse fra Dr. ing. Jan Roger Hoff, MARINTEK, Trondheim, 1998.

<sup>45</sup> Ibid.

<sup>46</sup> *Norsk Lovtidend*, avd I. nr 6(2002).

ekstern vurdering av evalueringsordningene. I merknadene blir det likevel foreslått at det bør være en viss ekstern deltakelse ved tradisjonell sensur både for høyere og lavere grad.

Hensikten med ekstern evaluering av vurderingen er også å ivareta studentenes rettssikkerhet ved å sikre lik og upartisk bedømmelse av studentene.<sup>47</sup> Merknadene er imidlertid ikke utfyllende i forhold til hvordan ekstern deltakelse i vurderingen skal eller kan gjennomføres. Eksempler er ekstern deltakelse ved oppgaveformulering, fastsetting av vurderingskriterier og ekstern stikkprøvekontroll av sensur foretatt av intern sensor, eller ekstern sensur av bestemte andeler av vurderinger som inngår i endelig karakter. Det kan også være ekstern sensur av større skriftlige arbeid som prosjekter og porteføljer.<sup>48</sup> For å sikre kandidatens rettssikkerhet, skal det være ekstern sensur ved bedømmelse av selvstendig arbeid ved høyere grad.<sup>49</sup>

En student kan klage skriftlig på karakteren for egen prestasjon innen tre uker etter at eksamensresultatet er gjort kjent. Det skal da benyttes to sensorer, hvorav minst en ekstern. Endringer kan gjøres både til gunst og ugunst for den som klager. Institusjonen kan selv bestemme om hver enkelt delprøve kan påklages, eller om klage først kan fremmes når alle delkarakterer som inngår i grunnlaget for endelig karakter er kunngjort. Vurderingsuttrykket ved eksamen eller prøve skal være bestått eller ikke bestått, eller en gradert skala med fem trinn fra A til E for bestått, og F for ikke bestått. Dermed har vi fått en karakterskala som er i tråd med ECTS (European Credit Transfer System). Hensikten er blant annet å kunne sammenlikne og kombinere norsk og internasjonal utdanning. Dette innebærer også en ny vekting, slik at et studieår gir en uttelling på 60 studiepoeng i stedet for dagens 20 vekttall.

## 11. Oppsummering og konklusjon

Begrepet “nye vurderingsformer” er i noen tilfelle lite dekkende i forhold til den virkelighet vi forsøker å beskrive. I konkrete sammenhenger er det ikke nødvendigvis formen som er ny, men derimot sammenstillingen og bruken av ulike former for bestemte formål. Hvilken eller hvilke former som blir valgt, er avhengig av hvilke funksjoner som ønskes ivaretatt. Mens tradisjonell slutteksamen i hovedsak har hatt en *summativ* funksjon, har de i den senere tid blitt stadig større oppmerksomhet rundt *formative* aspekter ved vurdering. Det betyr

---

<sup>47</sup> UFD – Ot. prp. Nr. 40 (2001-2002:12).

<sup>48</sup> Ibid.

<sup>49</sup> Op. cit.:13.

at vurdering blir et viktig redskap underveis i selve læringsarbeidet, ikke bare en kontroll av resultatene av det. Jeg har i den anledning benyttet en "ramfaktorteoretisk" tilnærming, der jeg ser vurderingsordningene som grunnleggende for studentenes handlingsrom. Valg av nye eller alternative vurderingsformer handler om å sannsynliggjøre hensiktsmessig studieatferd.

Selv om det finnes fellestrekk, er alle undervisnings- og læringssituasjoner unike. Ikke bare er fag og læringsmål ulike, det er også studentene og faglærernes forutsetninger. Læring skjer med andre ord under ulike betingelser. Diskusjonen om "nye" vurderingsformer handler ikke bare om å erstatte en type vurdering med en annen, som ansees som ubrukelig. Kunnskapskontroll, rangering og sertifisering er fortsatt en fullt legitim og sentral funksjon ved siden av en noe større vekt på lærings- og utviklingsmulighetene ved vurdering. Å få til dette handler ikke bare om pedagogisk og didaktisk innsikt. Det er ressurs-situasjonen samlet (tid, rom, penger) som setter rammer for hva som er mulig å få til, også der endringer er forventet og ønsket.

## Bibliografi

- Assessment Principles*, University of Queensland, Australia. URL <http://www.uq.edu.au/hupp/contents/view.asp?s1=3&s2=30&s3=1> <05.02.03>
- Berglind, H. (1995): *Handlingsteori och mänskliga relationer*. Borås: Natur och Kultur.
- Biggs, J. B. (s.a.): "Teaching Design for Learning". Keynote Paper, Brisbane: Higher Education Research and Development Society of Australia.
- Boud, D. (1995): *Enhancing Learning Through Self Assessment*. London: Kogan Page.
- Boudon, R. (1981): *The logic of social action. An introduction to sociological analysis*. London: Routledge & Kegan Paul.
- Dahllöf, U (1997): En återblick på det tidiga ramfaktorteoretiske tänkandet och några av dess senare utlöpare. Bidrag til symposiet *På återbesök i ramfaktorteorin*. Pedagogiska institutionen, Uppsala universitet, 20.-21.03.
- Gibbs, G. (1992): *Improving the Quality of Student Learning*. Bristol: Technical and Educational Services, Ltd.
- Gyldendals fremmedordbok*. Oslo: Gyldendal. 1965.
- Gynnild, V. (2001): *Læringsorientert eller eksamensfokuset? Nærstudier av pedagogisk utviklingsarbeid i sivilingeniørstudiet*. Trondheim: Fakultet for samfunnsvitenskap og teknologiledelse, Pedagogisk institutt, Norges teknisk-naturvitenskapelige universitet.
- Harris, D. & Bell, C. (1990): *Evaluating and Assessing for Learning*. London: Kogan Page.
- James, R., McInnis, C. & Devlin, M. (2002): *Assessing Learning in Australian Universities*. Centre for the Study of Higher Education (CSHE). The University of Melbourne.
- Kvale, S. (1970): *En eksaminasjon av universitetseksamener*. Oslo: Universitetsforlaget.
- Kvalsund, R. & Bøe, L. & Heggen, K. & Madssen, K.A. & Vaage, S. (red.) (1998) : *I lærande fellesskap. Helsingsskrift til Urban Dahllöf på 70-årsdagen*. Høgskolen i Volda/ Møreforskning.

- Marton, F. & Säljö, R. (1976): On Qualitative Differences in Learning. I: Outcome and Process. *British Journal of Educational Psychology*, 46(1976).
- Mowl, G. (s.a.): *Innovative student assessment: what's the point?* University of Northumbria. *Norsk Lovtidend I*, nr 6(2002).
- Rowntree, D. (1987): *Assessing Students: How shall we know them?* London: Kogan Page.
- Snyder, B. R. (1971): *The Hidden Curriculum*. New York: Knopf.
- St. meld nr. 27 (2000-2001)*: "Gjør din plikt – Krev din rett. Kvalitetsreform av høyere utdanning." Oslo: Kirke-, utdannings- og forskningsdepartementet.
- Thematic review of the first years of tertiary education*. Country note: Norway. (OECD Publications). URL <http://www1.oecd.org/els/education/ep/docs.htm> <05.02.03>
- UFD – Ot. prp. Nr. 40 (2001-2002)*. "12 Merknader til de enkelte bestemmelsene"
- Von Wright, G. H. (1993): *Explanation and Understanding*. New York: Cornell University Press.