

# Kriteriebaserede evalueringer

2004

# Kriteriebaserede evalueringer

2004

EVA's erfaringer fra evalueringer baseret på kriterier

## Kriteriebaserede evalueringer

© 2004 Danmarks Evalueringsinstitut  
Trykt hos Vester Kopi

Eftertryk med kildeangivelse er tilladt

Bestilles hos:

Danmarks Evalueringsinstitut  
Østbanegade 55.3  
2100 København Ø

T 35 55 01 01  
F 35 55 10 11

E [eva@eva.dk](mailto:eva@eva.dk)  
H [www.eva.dk](http://www.eva.dk)

|          |                            |          |
|----------|----------------------------|----------|
| <b>1</b> | <b>Indledning</b>          | <b>3</b> |
| <b>2</b> | <b>Konklusioner</b>        | <b>4</b> |
| <b>3</b> | <b>Begrebsafklaring</b>    | <b>5</b> |
| 3.1      | Kriterier                  | 5        |
| 3.2      | Indikatorer                | 6        |
| 3.3      | Et eksempel                | 6        |
| <b>4</b> | <b>Erfaringer</b>          | <b>7</b> |
| 4.1      | Grundlaget for kriterierne | 7        |
| 4.2      | Gennemsigtighed            | 7        |
| 4.3      | Sammenlignelighed          | 8        |
| 4.4      | Involvering                | 8        |
| 4.5      | Kontrol og udvikling       | 8        |
| 4.6      | Antallet af kriterier      | 9        |
| 4.7      | Begrebsafklaring           | 9        |
| 4.8      | Afsluttende bemærkninger   | 9        |



EVA har inden for de sidste to år høstet sine første erfaringer med at gennemføre evalueringer der anvender eksplicitte kriterier som vurderingsgrundlag. Først og fremmest drejer det sig om evalueringen af tre samfundsvidenskabelige masteruddannelser og en international evaluering af agronomuddannelsen i fire europæiske lande. Begge disse evalueringer havde karakter af pilotprojekter hvor EVA lagde vægt på at indsamle metodiske erfaringer, blandt andet med hvilke styrker, svagheder og muligheder denne type evalueringer byder på. Siden hen har EVA gennemført flere kriteriebaserede evalueringer. Det drejer sig om en evaluering af de klassiske konservatorieuddannelser og en transnational europæisk evaluering af fysik. Sidstnævnte evaluering er først afsluttet i 2003. Sidst, men ikke mindst har EVA for Undervisningsministeriets udarbejdet kriterier for professionsbacheloruddannelserne med det formål at præcisere ministeriets forventninger til tildelingen af en professionsbachelorortitel.

EVA's interesse for at indsamle metodiske erfaringer med kriteriebaserede evalueringer er især motiveret af den internationale udvikling hvor brugen af kriterier i evalueringer i løbet af de seneste år har vundet stigende indpas. Det hænger blandt andet sammen med betoningen af behovene for åbenhed og gennemsigtighed som led i internationaliseringen af det videregående uddannelsesområde i tilknytning til implementeringen af Bologna-processen.

På den baggrund og set i lyset af at denne evalueringstype fortsat er relativt ny i EVA-sammenhæng, eksisterer der et betydeligt behov for at opsamle og systematisere de foreløbige erfaringer med disse evalueringer.

### **Formål og indhold**

Formålet med denne publikation er således at samle EVA's erfaringer med kriteriebaserede evalueringer for på den baggrund at kvalificere brugen af kriterier i fremtidige evalueringer. Desuden skal publikationen afklare EVA's brug af kriteriebegrebet og medvirke til at erfaringerne bliver udbredt.

Publikationen vil blandt andet belyse følgende spørgsmål:

- Hvad er værditilvæksten ved kriteriebaserede evalueringer sammenlignet med evalueringer der ikke anvender eksplicitte kriterier? Hvilke nye metodeelementer tilfører denne evalueringstype EVA's evalueringer, og i hvilken udstrækning er den forenelig med en fitness for purpose-tilgang, dvs. en tilgang der tager udgangspunkt i uddannelsernes egne definerede mål og principper?
- Hvem fastsætter kriterierne, og hvordan sikrer man forankring hos evalueringens centrale aktører?
- Hvordan undgår man at kriteriebaserede evalueringer fører til ensretning og begrænset nyudvikling af de evaluerede uddannelser?
- Hvordan sikrer man udviklingsaspektet bedst muligt i evalueringerne?

Trods udfordringerne har der været en klar værditilvækst ved at bruge kriterier som vurderingsgrundlag i pilotevalueringerne. Særlig vigtigt er følgende:

- **Åbenhed og gennemsigtighed**

Kriterier giver øget åbenhed og gennemsigtighed i evalueringen gennem ekspliciteringen af vurderingsgrundlaget og værditilvæksten. Det er klart på hvilke vilkår evalueringen gennemføres, og grundlaget for identifikation af god kvalitet er åbent og kendt. Man kan sige at kriterierne udgør evalueringens kvalitetsbegreb.

- **Fokusering**

Kriterier giver god fokusering i evalueringen og en bedre mulighed for løbende at fastholde fokus indadtil i forhold til evalueringsgruppen og udadtil i forhold til de evaluerede uddannelsesinstitutioner. En fokuseret evaluering kræver dog at der i forbindelse med formuleringen og valget af kriterier foretages en forholdsvis hård prioritering mellem vigtigt og mindre vigtigt så antallet af kriterier holdes på et fornuftigt niveau. Resultatet af en sådan fokusering skulle gerne være at kriterierne tilsammen dækker bredere, og at de ikke overlapper og gentager hinanden.

- **Samarbejde og dialog**

Kriterier udgør et godt og konstruktivt grundlag for samarbejde mellem EVA og evalueringsgrupper og for dialog med uddannelsesinstitutionerne. Ved at inddrage evalueringsgrupperne og uddannelsesinstitutionerne i formuleringen og udvælgelsen af kriterier kan evalueringen sikres en høj faglig præcision, en oplevelse af medejerskab og dermed en høj grad af legitimitet. Samtidig gør EVA's særlige erfaring og kompetence i forhold til at formulere kriterier det vigtigt at instituttet fastholder sin position som garant for de endelige kriteriers soliditet.

- **Inddragelse af standarder**

Kriterier udgør et godt grundlag for at sikre at evalueringen inddrager og balancerer alle de relevante standarder. Erfaringerne viser at kriterierne både kan rumme fx lovkrav, mere kvalitative aspekter som evalueringsgruppens krav og forventninger, og uddannelsesinstitutionernes egne mål og principper.

- **Kontrol og udvikling**

Kriterier giver gode muligheder for at kombinere kontrol- og udviklingshensyn i evalueringen ved at anvende en blanding af forskellige typer af kriterier, fx minimums- og best practice-kriterier, fitness for purpose-baserede kriterier og åbne og lukkede kriterier.

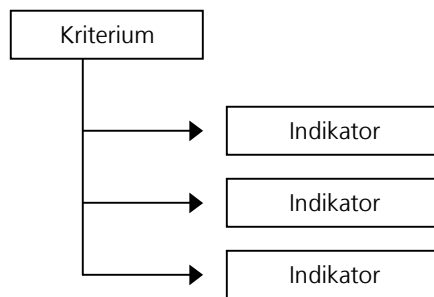
- **Afrapportering**

Kriterier muliggør en forholdsvis kortfattet afrapportering hvor man kan fokusere på de kriterier som ikke er opfyldt.

Desuden ligger der en udfordring i at afbalancere en tilbøjelighed til at fokusere på de enkelte kriterier på bekostning af helhedsperspektivet.

Kriterier er som begreb hverken nationalt eller internationalt entydigt i forbindelse med evaluering, og det blandes ofte sammen med begreber som standarder, benchmarks og indikatorer. Især bruges kriterier i mange sammenhænge synonymt med standarder. For at klargøre betydningen af de enkelte begreber og deres indbyrdes sammenhæng giver dette kapitel et bud på definitioner. Definitionerne er rettet mod at indfange anvendelsen af begreberne, snarere end at give en mere teoretisk forankret definition. Det ses blandt andet ved at definitionen af hhv. standarder og kriterier ikke er gensidigt udelukkende, men til gengæld indkredser forskellige funktioner, detaljeringsniveauer mv. som er relevante i forhold til praktisk brug og udfoldelse af begreberne.

Som det vil fremgå, ser EVA en hierarkisk relation mellem begreberne hvor kriterier er mere overordnede end indikatorer. Begreberne forholder sig til hinanden på følgende måde:



### 3.1 Kriterier

*“Kriterier er operationaliserbare kvalitetskrav.”*

EVA definerer *kriterier* som en betegnelse for operationaliserbare udsagn om kvalitet i undervisning og uddannelse. Kriterier er altså fokuserede kvalitetskrav der er tæt på det operationelle niveau uden at de dog nødvendigvis indeholder en specifik målestok.

I forhold til EVA’s evalueringer er det nye at der i forbindelse med evalueringer kan blive formuleret en række eksplicite krav som den evaluerede uddannelse eller uddannelsesinstitution skal leve op til for at der kan tales om god kvalitet. I modsætning til EVA’s traditionelle fitness for purpose-evalueringer indebærer kriteriebaserede evalueringer derfor foruddefinerede udsagn om hvad acceptabel eller særlig god kvalitet er.

#### **Fitness for purpose og kriterier**

Fitness for purpose betegner evalueringer hvor grundlaget for vurdering af kvaliteten er forankret i uddannelsens egne formål og principper. Der er dog ikke nødvendigvis et modsætningsforhold mellem fitness for purpose og brugen af kriterier som grundlag for evaluering. Flere af EVA’s kriteriebaserede evalueringer har indeholdt fitness for purpose-elementer ved at kriterierne var meget rummelige og åbne for de enkelte institutioners fortolkninger. Det har eksempelvis været tilfældet når kriterierne knyttede an til de pågældende institutioners eller uddannelsers målsætninger.

#### **Kriterier som grundlag for vurdering**

Kriterier skal imidlertid udtrykke kvalitet så præcist at der sikres et gennemsigtigt grundlag for vurdering. Kriterier er således testbare udsagn der er formuleret på baggrund af accepterede udsagn med henblik på at undersøge om det ønskede kvalitetsniveau er nået. Der kan og bør være en tæt sammenhæng mellem kriterier og udsagn.

### Minimumskrav og best practice

Kriterier kan både udtrykke minimumskrav og best practice. Ved minimumskrav forstås kriterier som beskriver forhold der skal være opfyldt for at kvaliteten er acceptabel – typisk lovgivningsmæssige forhold. Ved best practice forstås krav som skal opfyldes for at uddannelsesaktiviteten kan siges at være på et særlig godt kvalitetsniveau inden for et bestemt område.

### Benchmarks

En særlig type kriterier er benchmarks. Benchmarks er kriterier som er forankret i sammenligning med andres praksis. De bruges som eksplicite reference- eller målepunkter i relation til systematiske sammenligninger der skal bruges til at afdække et forbedrings- eller effektiviseringspotentiale for de involverede uddannelser eller institutioner. Benchmarks er altid forankret i analyser af best practice eller særlig god kvalitet med henblik på at synliggøre de involverede uddannelsers eller institutioners forbedringspotentiale.

## 3.2 Indikatorer

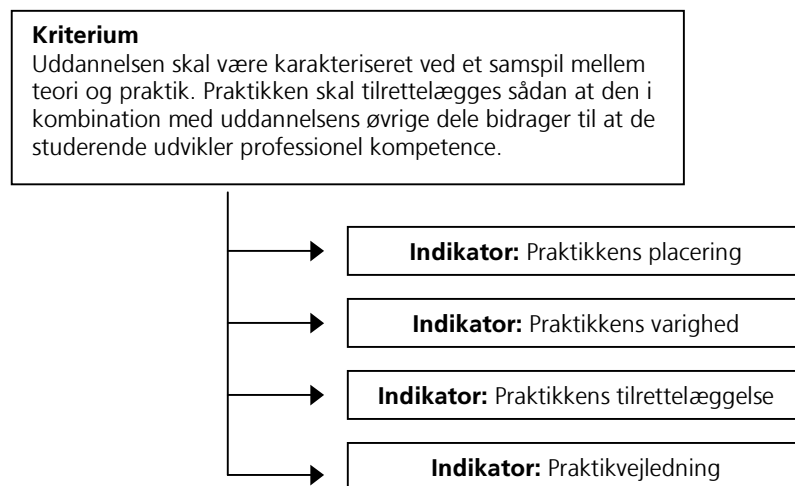
*“Indikatorer er konkrete mål for kriterier.”*

EVA definerer *indikatorer* som konkrete, målbare, kvalitative eller kvantitative variable der udtrykker centrale aspekter ved den institution, uddannelse eller undervisning som ønskes vurderet gennem anvendelsen af kriterier. Indikatorerne skal tilsammen tegne et dækkende billede af om uddannelsen opfylder de anvendte kriterier.

Der anvendes typisk flere indikatorer til at give en tilfredsstillende belysning af et kriterium. Udarbejdelsen af indikatorer sker gennem en operationalisering af kriterierne på en sådan måde at der sikres en høj grad af validitet gennem tæt sammenhæng mellem indikatorer og kriterier. Indikatorer relaterer sig til alle aspekter af uddannelse og kan både give direkte og indirekte beskrivelser.

## 3.3 Et eksempel

For at tydeliggøre begreberne følger her et konstrueret eksempel på hvordan kriterier og indikatorer i praksis kunne være formuleret.



Som det ses, giver indikatorerne tilsammen det informationsgrundlag som er nødvendigt for at kunne vurdere om kriteriets krav er opfyldt.

De evalueringer som dette notat bygger på, indeholder en række fælles træk som afspejler den løbende erfaringsudveksling der er foregået mellem projekterne. Imidlertid illustrerer eksemplerne også en betydelig variation i tilgang og indhold. Denne variation er et udtryk for at EVA har tilstræbt at afprøve forskellige modeller og tilgange. Forskelligheden bør også fremover kunne rummes i EVA's kriteriebaserede evalueringer under hensyntagen til evalueringernes forskellige formål og fokus.

Dette kapitel er en opsamling på EVA's væsentligste erfaringer i forbindelse med kriteriebaserede pilotevalueringer.

## 4.1 Grundlaget for kriterierne

Kriterierne i pilotevalueringerne udspringer af en bred vifte af kilder. I masterevalueringen blev kriterierne formuleret på baggrund af relevant lovgivning og eksisterende undersøgelser på området, erfaringer fra Evalueringscenteret, evalueringsgruppens og institutionernes faglige indsigt og erfaringer. De kriterier der blev anvendt i agronomevalueringen, bygger på et lignende grundlag, dog spillede evalueringsgruppen ikke den samme rolle. Desuden havde de lovgivningsmæssige rammer ikke den samme betydning på grund af evalueringens internationale karakter og forskelle i de nationale lovgivninger på området. Det samme gjorde sig gældende i forhold til musikkonservatorieevalueringen der overvejende hentede inspiration fra øvrige kriteriebaserede evalueringer, litteratur- og artikelstudier samt evalueringsgruppens faglige indsigt i området.

Professionsbachelorkriterierne er helt overvejende juridisk begrundet i professionsbachelorbekendtgørelsen, og Undervisningsministeriet har vurderet kriteriernes juridiske holdbarhed.

I samtlige pilotevalueringer blev nedenstående krav desuden anvendt som retningslinjer for formuleringen af kriterier. Evalueringernes kriterier skulle:

- være relevante for det pågældende uddannelsesniveau
- være ensartede og have bred gyldighed
- være rummelige og indeholde et fitness for purpose-element
- være operationelle og danne grundlag for reel vurdering
- fokusere på udmøntning og ikke blot principper
- være konsistente i forhold til hinanden.

## 4.2 Gennemsigthed

Erfaringen fra samtlige pilotprojekter peger på at kriterier som evalueringsmetode har sin væsentlige styrke i at kriterierne skaber gennemsigthed i vurderingsgrundlaget. I forhold til pilotevalueringerne har det betydet at evalueringsgruppen har haft en fælles referenceramme for hvad der opfattes som "good practice". Det gælder især for master- og musikkonservatorieevalueringen hvor evalueringsgruppen var involveret i udformningen af kriterierne, og hvor kriterierne derfor synliggjorde evalueringsgruppens kvalitetsforståelse. Kriterierne udtrykte åbent hvad der ifølge evalueringsgruppen er god kvalitet.

For uddannelsesinstitutionerne har brugen af kriterier tydeliggjort hvilke præmisser der er lagt til grund for vurderingen af deres styrker og svagheder og for evalueringens anbefalinger. For EVA har brugen af kriterier vist sig som et effektivt redskab til at fastholde et relevant fokus i evalueringerne.

### 4.3 Sammenlignelighed

Som metode indebærer kriterier også den fordel at de er velegnede som fikspunkt for systematiske sammenligninger. Det har især vist sig i forhold til musikkonservatorieevalueringen og agronomevalueringen der begge sammenlignede uddannelsernes kvalitet på tværs af institutioner og landegrænser. Samtidig har det dog vist sig at andre hensyn som for eksempel integrering af et fitness for purpose-element kan besværliggøre sammenligninger.

### 4.4 Involvering

Erfaringerne fra pilotevalueringerne er at både projekt- og evalueringsgrupperne spiller centrale roller i forbindelse med formuleringen af kriterier.

I både masterevalueringen og musikkonservatorieevalueringen deltog evalueringsgruppen i formuleringen af kriterierne. Kriterierne forpligtede under evalueringsprocessen evalueringsgruppens medlemmer på det valgte fokus og kriterier og var velegnede til at strukturere diskussionerne af de forskellige aspekter af uddannelserne.

Dialogen med evalueringsgrupperne om kriterierne er vigtig af hensyn til den faglige kvalificering og sikringen af medejerskab, men samtidig har EVA en særlig erfaring og kompetence i forhold til at formulere kriterier. Det er derfor vigtigt at EVA fastholder sin position som garant for de endelige kriteriers soliditet.

Institutionerne har i varierende omfang været inddraget i formuleringen af kriterierne for de tre evalueringer. Erfaringen fra pilotevalueringerne viser desuden at involveringen af institutionerne har bidraget til at skabe legitimitet og medejerskab om den eksterne vurdering og sikret et tillidsfuldt udgangspunkt for den videre proces.

Erfaringerne fra agronomevalueringen var på dette område anderledes. I lighed med masterevalueringen var erfaringen herfra at involveringen af institutionerne i nogen grad bidrog til at sikre at kriterierne har været dækkende og relevante i forhold til de evaluerede uddannelser. Desuden betød involveringen også at kriterierne havde en praksisnærhed og derfor ikke kun udtrykte hensigtserklæringer og principper. Involvering af institutionerne skal imidlertid balanceres i forhold til hensynet om at sikre uafhængighed i vurderingsgrundlaget.

### 4.5 Kontrol og udvikling

Den generelle erfaring fra de fire projekter er at udviklingsperspektivet bedst sikres ved at der indgår åbne kriterier fordi de i højere grad end lukkede lægger op til refleksion på uddannelser og institutioner. Dette skal dog ses i sammenhæng med valget mellem minimumskriterier og best practice-kriterier. Når kriterierne er knyttet til et best practice-perspektiv, er det med til tydeligt at rette fokus på muligheder og behov for kvalitetsudvikling. Et åbent minimumskriterium kan således lægge op til refleksion om forskellige måder at leve op til kriteriet på, men vil typisk ikke have så vidtgående udviklingsperspektiver som best practice-kriterier. Det er dog i den sammenhæng vigtigt at være opmærksom på at kriterierne skal relatere sig til "virkelige" eksempler der gør at kriterierne ikke er urealistiske at efterleve, og at institutionerne ikke måles i forhold til abstrakte målsætninger. I den sammenhæng kan det overvejes at benytte good practice frem for best practice da selv de bedste institutioner eller uddannelser kan præstere resultater der ikke er optimale i absolut forstand. Minimumskriterier er hensigtsmæssige at anvende i det omfang man ønsker at anvende kriterierne til at udtrykke krav og betingelser der er knyttet til relevant lovgivning på uddannelsesområderne.

De fire projekter har alle anvendt en del kriterier som er fitness for purpose-baserede. Erfaringen med at integrere fitness for purpose-elementet er at det er velegnet til at sikre mod at kriterierne fører til ensretning og begrænset nyudvikling af uddannelserne. Desuden er fitness for purpose-elementet med til at sikre den rummelighed i vurderingerne som EVA tilstræber. Imidlertid kan elementet gøre at uddannelserne ikke uden videre kan sammenlignes på grund af forskelle i uddannelsernes formål og målsætninger. Det er især problematisk for evalueringer med et stærkt komparativt sigte. Det er vanskeligt entydigt at pege på hvilke af de to hensyn der vejer tungest. Det må afhænge af den konkrete sammenhæng, blandt andet evalueringens formål.

## 4.6 Antallet af kriterier

Pilotevalueringerne viste at der findes nogle centrale udfordringer i at afbalancere kriteriebaserede evalueringer med hensyn til antallet af kriterier og de enkelte kriteriers kompleksitet, altså hvor mange dimensioner de enkelte kriterier indeholder.

Den generelle erfaring er at antallet af kriterier var meget omfattende, og at der derfor i nogen grad også forekom overlap og gentagelser. Set i bakspejlet kunne der have været foretaget en sanering af kriterierne som kunne have bidraget til at undgå gentagelser og skelne væsentligt fra mindre væsentligt. Omfanget af kriterier blev også fra institutionernes side fremhævet som (unødigt) ressourcekrævende i forhold til selvevalueringsprocessen.

I musikkonservatorieevalueringen og agronomevalueringen var det desuden erfaringen at de enkelte kriterier i praksis har vist sig vanskelige at adskille. Det gælder især de kriterier der handler om henholdsvis mål og indhold.

I forbindelse med masterevalueringen blev det fremhævet at nogle af kriterierne var for sammensatte og indeholdt for komplekse relationer. Det medførte at kriteriernes dynamik i nogen grad blev formindsket. Omvendt må det dog understreges at det er vigtigt at kriterierne ikke bliver så simple at de udtrykker et for enkelt kvalitetsbegreb og dermed ikke modsvarer kompleksiteten i de uddannelsesmæssige forhold som evalueres.

## 4.7 Begrebsafklaring

En væsentlig erfaring fra agronomevalueringen var betydningen af at udarbejde præcise definitioner af de centrale begreber der optræder i kriterierne, eksempelvis "professional qualifications". På den måde minimerer man risikoen for misforståelser og misfortolkninger i institutionernes selvevalueringsrapporter. Det er især vigtigt i en international evaluering hvor de selvevaluerende institutioner i mange tilfælde opererer på engelsk. Forskellige udlægninger af begreberne kan føre til uensartede svar der kan være vanskelige at sammenligne på tværs. I fremtidige evalueringer bør man derfor overveje at udarbejde et "glossary" med ordforklaringer der vedlægges selvevalueringsguiden.

## 4.8 Afsluttende bemærkninger

EVA's første erfaringer med kriteriebaserede evalueringer har været af stor værdi for de nye kriteriebaserede projekter som instituttet for nylig har påbegyndt og fremover vil overveje at tage initiativ til. Parallelt med EVA's egne erfaringer og refleksioner herover vil instituttet løbende følge de erfaringer andre evalueringsorganisationer gør sig på området.