

Danish University Colleges

Notat om tværgående analyse af vurderingskriterier

Lund, Jens H; Buch, Bettina; Revier, Robert Lee

Publication date:
2020

[Link to publication](#)

Citation for pulished version (APA):

Lund, J. H., Buch, B., & Revier, R. L. (2020). *Notat om tværgående analyse af vurderingskriterier.*

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Download policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Notat om tværgående læsning af vurderingskriterier (sept. 2020).

V. Bettina Buch (UC Absalon), Jens H. Lund (VIA) og Robert Lee Revier (UCL)

En arbejdsgruppe bestående af Bettina Buch (UC Absalon), Robert Lee Revier (UCL) og Jens H. Lund (VIA) har gennem en kort periode på fire uger gennemlæst de indsendte forslag til vurderingskriterier fra samtlige fag på læreruddannelsen. I dette notat indleder vi med at angive grundlaget for opgaveløsningen, hvorefter vi redegør for vores metodiske tilgang til opgaveløsningen. Her opstiller vi også vores kriterier for vores vurdering. Med afsæt i den skitserede vurderingsramme besvarer vi til sidst de fire spørgsmål, som er angivet i kommissoriet for opgaven.

Baggrund

Arbejdsgruppen er ikke nærmere bekendt med de analyser og problemdefinitioner, som ligger bag LLNs udpegning af vurderingskriterier som løsning på de problemstillinger, der nævnes i kommissoriet for opgaven:

LLN (har) sammen med de nationale faggrupper og censorforpersoner på møder i 2019/20 drøftet prøver og kriterier for bedømmelse i lyset af de seneste censorrapporter, det meget høje karaktergennemsnit i en række fag og ekspertgruppens evaluering af læreruddannelsen, der bl.a. kritiserer karakterniveauet i BA-prøverne. Målet med at udarbejde vurderingskriterier er at formulere kriterier, som kan danne grundlag for vurdering, der samtidig skærper kravene til de studerende i forhold til fagligt niveau i fagene.
(Opgavekommissorium, 2020)

Vi forholder os heller ikke nærmere hertil i dette notat, men vil dog ikke undlade at referere, at vi i arbejdsgruppen har drøftet den begrænsede forandringskraft, der kan ligge i *udelukkende* at formulere vurderingskriterier. Vi formoder, at der kan udpeges en bred vifte af tiltag, som kan bidrage til løsning af de nævnte problemstillinger. Vi er endvidere af den opfattelse, at man kan forholde sig kritisk/problematiserende til de angivne problemdefinitioner. Som det vil fremgå af nedenstående, refererer vi ikke disse drøftelser yderligere, men fokuserer i stedet at besvare de fire spørgsmål i kommissoriet ud fra de angivne præmisser og rammer.

Jævnfør rammerne for opgaveløsningen er det begrænset, hvad vi har kunnet formå at inddrage af teoretisk og forskningsmæssig grundlag. Men via velvillig sparring fra evalueringseksperter Jakob Wandall (se metode nedenfor) er vi blevet gjort opmærksom på en central og inspirerende artikel af Johnsson & Svingby, (2007), som vi derfor kort refererer i det følgende.

Johnsson & Svingby peger på, at når man indfører vurderingskriterier er baggrunden, at man ønsker en sikker og ensartet, retfærdig bedømmelse. Den ideelle situation er, at enhver præstation altid bedømmes ens af alle bedømmere og at den også bedømmes ens fra gang til gang:

Ideally, an assessment should be independent of who does the scoring and the results similar no matter when and where the assessment is carried out, but this is hardly obtainable (Johnsson & Svingby, 2007: 133).

De påpeger dog også, at sikkerheden for elever og studerende er svær at garantere, hvis ikke alle stilles den samme opgave, idet dette yderligere skaber vanskeligheder ved at gennemføre en ensartet bedømmelse:

Of course, performance assessment is not traditional testing, and several findings in this review support the rather self-evident fact that when all students do the same task or test, and the scoring procedures are well defined; the reliability will most likely be high. But when students do different tasks, choose their own topics or produce unique items, then reliability could be expected to be relatively low (Brennan, 1996) (Johnsson & Svingby, 2007).

Prøver hvor de studerende ikke gennemfører nøjagtig den samme prøve, er almindelige på læreruddannelsen i Danmark, og en valid og reliabel vurdering er således svær at sikre. Vurderingskriterier kan være én vej at gå.

Ifølge Johnsson og Svingby (2007) er man gået fra at gennemføre det, de kalder traditionelle test, som primært er multiple choice-test til det, de kalder 'performance assessment'. Dette gælder nok ikke helt for Danmark, hvor multiple choice-test aldrig har været en foretrukket og almindelig testform. Snarere er netop multiple choice-opgaver noget, der også er kommet frem i de senere år, hvor fokus har været på testning i det hele taget.

Johnsson og Svingby finder, på baggrund af Messicks model for testning, at 'performance assessment' ligger i den ene ende af et kontinuum, mens multiple choice-test ligger i den anden. De betegner det første som den bløde ende. Ser man på testning på den måde, har vi i Danmark i lang tid haft en tradition for prøver og test, der ligger i den 'bløde' ende af dette kontinuum. Performance assessment betegnes som prøver, der lægger sig op ad aktiviteter, der kan forekomme i den 'virkelige verden', mere eller mindre direkte som en slags efterligning af reelle problemstillinger (Johnsson & Svingby, 2007). Denne type prøver er svære at gennemføre på en valid og reliabel måde samtidig med, at prøvers større indflydelse på elevens og studerendes fremtid medfører, at kravet til netop validitet og reliabilitet i vurderingen af præstationer øges. Vurderingskriterier, også kaldet rubrics på engelsk, kan understøtte større validitet og især reliabilitet.

Endelig skelner Johnsson og Svingby mellem en holistisk og en analytisk scoring af præstationer, hvor den holistiske scoring betegner en metode, hvor der angives en samlet score på hele præstationen, mens en analytisk scoring består af delscorer på hver opgavedel. Her anfører de, at den analytiske scoring er mere sikker, dvs. valid og reliabel, end den holistiske. Også her kan vurderingskriterier altså understøtte karaktergivning, da de kan konstrueres, så de understøtter en analytisk vurdering gennem kriterier for en række delområder.

Johnsson og Svingby anbefaler en række tiltag, der kan styrke validitet og reliabilitet. Reliabiliteten styrkes ved, at vurderingskriterierne giver mulighed for en analytisk bedømmelse, dvs. der skal scores/vurderes på flere inkluderede delparametre, idet

muligheder for at nuancere karakteren kan styrke ligheden mellem to bedømmere (inter-rater-reliabilitet), de skal være fagspecifikke, eller egentlig emnespecifikke, og der skal gives konkrete eksempler (Jonsson & Svingby, 2007: 131). Desuden vil det være en styrke at træne bedømmere for at opnå større enighed blandt dem. Dvs. at man samles om at bedømme og diskutere konkrete præstationer i faggrupper. Endelig konkluderes det, at vurderingskriterier ikke sikrer validiteten automatisk. Her kræves det, at kriterierne afprøves empirisk. Der er også empirisk belæg for, at brugen af eksplicite vurderingskriterier kan styrke læringen, da de kan understøtte feedback og selvvurderinger.

Metodisk tilgang

Arbejdsgruppen har gennemgået en iterativ proces, der har bestået af følgende elementer

- Indledende eksplorative analyser af ni udvalgte forslag fra de nationale faggrupper (NT., Idræt, Engelsk, Praktik, Matematik, Dansk, Musik og PL)
- sparring med ekspert: Jakob Wandall, der bl.a. er medforfatter til bogen: Andersen, M., Wahlgren, B. & Wandall, J. (2017). *Evaluering af læring, undervisning og uddannelse*. Hans Reitzels Forlag
- litteraturstudier, bl.a. : Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2), 130-144. link: [https://www.homeworkgain.com/wp-content/uploads/edd/2019/10/20190720223824jonsson_svingby_2007 .pdf](https://www.homeworkgain.com/wp-content/uploads/edd/2019/10/20190720223824jonsson_svingby_2007.pdf)
- [og: Brookhart, S.M. \(2013\). How to create and use rubrics for formative assessment and grading. Oxford Handbooks for language teachers.](#)
- Udveksling af iagttagelser i arbejdsgruppen
- gennemførelse af analyse af de resterende bidrag fra de nationale faggrupper under anvendelse af udviklet analysemodel (se nedenfor).

Udviklet analysemodel

Som nævnt i metodeafsnittet har arbejdsgruppen udviklet en analyseramme, som er blevet anvendt i læsning og analyse af de fremsendte forslag fra de nationale faggrupper. Analyserammen kan samtidigt anskues som pejlemærker for det fremtidige arbejde med at justere og videreudvikle vurderingskriterier for de enkelte fag og deres kompetencemål.

Analyserammen rummer følgende fire kriterier for formulering af vurderingskriterier:

1. Iagttagelighed

Vurderingskriterierne beskriver iagttagelige fænomener, dvs. de studerendes tale, tekst og/eller gøren/performance

2. Niveauangivelser

Vurderingskriterierne angiver tydelige niveauer, hvilket indebærer dels, at der anvendes en eksplicit og forklaret taksonomi, dels at denne taksonomi anvendes

konsekvent. Arbejdsgruppen vil endvidere henlede opmærksomheden på § 21 i bekendtgørelse for læreruddannelsen. Heri angives en taksonomi, som kan kondenseres til flg.: redegøre, analysere, reflektere, vurdere, anvende, handle i henholdsvis kendte og ukendte situationer, udvikle praksis, selvstændighed. Det kan anbefales at det eksplicit angives, hvorledes en given taksonomi anvendes i vurderingskriterierne. To overordnede og forskellige tilgange er fx

1. om der anvendes en bred vifte af de forskellige niveauer i en taksonomi på hvert standpunktsniveau (standpunkt 2 indebærer fx at man både redegør og analyserer samt udvikler ny praksis) mm men at dette foregår i meget begrænset omfang, mens standpunktet 12 indebærer at de tilsvarende kompetencer udfoldes på højt niveau)
2. om taksonomien anvendes som en 'rækkefølge' og trinvis stigning (standpunktet 2 kan fx indebære, at man redegør, mens standpunktet 12 indebærer, at man selvstændigt udvikler ny praksis)

3. Faglig relevans

Vurderingskriterierne relaterer til formuleringerne i kompetencemålene, men er ikke gentagelse heraf.

4. Eksempler

Hvert vurderingskriterie er illustreret med et eksempel.

Besvarelse af de fire spørgsmål i kommissoriet

1) Hvad fungerer godt i de indsendte vurderingskriterier?

1.1 Når der er en tydelig bestræbelse på at angive iagttagelige fænomener (tale, tekst, performance) i vurderingskriterier - se fx Idræt og Tysk.

1.2 Når der er en tydelig bestræbelse på at angive og anvende en taksonomi i vurderingskriterierne, se fx Geografi.

1.3 Når der anvendes en analytisk vurdering, dvs. at der opdeles i mindre, faglige områder eller kompetencer. Dette har de fleste fag faktisk gjort.

2) Hvilke udfordringer er der i forhold til niveau og anvendelse

2.1 Generelt er de formulerede vurderingskriterier forholdsvist abstrakte og derfor vanskelige at anvende i en konkret vurderingssituation.

2.2 Flere faggrupper har valgt at kopiere standpunktskriterier ind som vurderingskriterier, hvorved der ikke formuleres vurderingskriterier, sådan som arbejdsgruppen har valgt at definere kategorien vurderingskriterier.

2.3 Flere faggrupper anvender en taksonomi - men taksonomien er ikke forklaret/eksplicit og det er heller ikke angivet, hvorledes taksonomien anvendes.

3) Hvilke anbefalinger kan på den baggrund gives ift det videre arbejde?

3.1 Det anbefales, at der udvikles en række kriterier for formulering af vurderingskriterier, som kan fungere retningsgivende og vejledende for det fortsatte arbejde i fx de nationale faggrupper. Det vil styrke såvel formuleringen af de enkelte vurderingskriterier som graden af konsistens i de samlede formuleringer.

3.2 Vi vurderer, at såvel udarbejdelse af kriterier for formulering af vurderingskriterier som anvendelse af disse i en evt. kommende revidering af de indsendte forslag til vurderingskriterier vil være ganske ressourcekrævende og måske kræver én eller anden form for yderligere støtte - fx konsulentstøtte, seminarer, videndeling etc. Johnsson og Svingby (2007) peger fx på potentialerne i at bedømmere arbejder tæt sammen om at bedømme de samme opgaver.

3.3 Derudover formoder vi, at en løsning af de angivne problemstillinger i kommissoriet kalder på en bred vifte af indsatser ud over arbejdet med formulering af vurderingskriterier.

3.4 Det kan anbefales, at de nationale faggrupper forholder sig til den taksonomi, der fremgår af bekendtgørelsens § 21 og at man derudover eksplicit forklarer, hvilken taksonomi der anvendes og hvordan den anvendes - se fx PL, som i en vis udstrækning har bevæget sig ud ad dette spor.

3.5 Det vil formodentlig være en hjælp i anvendelsen af vurderingskriterierne, at der angives et eksempel inden for hvert standpunkt. Som eksempel se fx. Musik, som har fyldige eksempler som bilag og NT, som har valgt at indsætte en case i hver af de tre standpunkter samt Geografi der også angiver helt konkrete eksempler.

4) Er der 3 eksemplariske eksempler på vurderingskriterier, som øvrige faggrupper kan lade sig inspirere af i forhold til kvalificering af vurderingskriterierne?

4.1 Vi ser os ikke i stand til at udpege ét fag eller flere som eksemplarisk men kan pege på enkelt eksempler fra forskellige fag, som tenderer at honorere de opstillede kriterier. Vi har derfor ovenfor valgt at give eksempler fra flere forskellige forslag til vurderingskriterier løbende gennem teksten.