

## CHAPTER 2

# Vurderingskriterier i utøvende musikk – et verktøy for pålitelig vurdering av studentenes bachelorkonsert?

*Wenche Waagen*

Norges teknisk-naturvitenskapelige universitet

**Abstract:** At the Norwegian University of Science and Technology (NTNU), Department of Music (IMU), new assessment guidelines were implemented during spring 2019, which had consequences for the assessment of bachelor concerts. The assessment guidelines offered two substantial new tools for external examiner use in summative assessment: They contained predefined assessment criteria and mark descriptions adapted to the expected learning outcome on the main instrument. In this chapter, I attempt to find out how the new tools function in practice and whether they clarify the task for the external examiners and increase transparency. A focus group interview with six examiners who used the guidelines at IMU is used to shed light on different aspects of the summative assessment. Sadler's theory of qualitative assessment and the twin concepts of holistic and analytic assessment are used as a basis for my discussion.

**Keywords:** bachelor concert, holistic and analytic assessment, mark descriptions, examiners, interview

I stortingsmeldingen *Kultur for kvalitet i høyere utdanning* (Meld. St. 16 (2016–2017)) varslet regjeringen at den vil stille krav om sensorveiledninger ved alle eksamener, med forhåndsdefinerte kriterier og

Sitering av dette kapitlet: Waagen, W. (2021). Vurderingskriterier i utøvende musikk – et verktøy for pålitelig vurdering av studentenes bachelorkonsert? I E. Angelo, J. Knigge, M. Sæther & W. Waagen (Red.), *Higher Education as Context for Music Pedagogy Research* (s. 41–64). Cappelen Damm Akademisk. <https://doi.org/10.23865/noasp.119.ch2>  
Lisens: CC BY-NC-ND 4.0.

karakterbeskrivelser. Sensorveiledninger presiserer, ut fra emnets læringsmål og innhold, hva studenten forventes å kunne. Hensikten er å bidra til en mer ensartet vurdering sensorene imellom, og å styrke studentens rettssikkerhet og kvalitet i utdanningene. Tidligere hadde UH-rådets<sup>1</sup> generelle, kvalitative beskrivelser av vurderingskriterier (UH-rådets utdanningsutvalg, 2016) ligget til grunn ved eksamener i utøvende musikkutdanninger. Rådet for utøvende musikkutdanning (RUM) nedsatte en arbeidsgruppe som utarbeidet felles og fagspesifikke forslag til sensorveiledninger, blant annet i hovedinstrumentfaget. Eksterne og interne sensorer har nå gjort erfaringer med disse. For å utdype min overgripende problemstilling, som er å undersøke hvorvidt forhåndsdefinerte vurderingskriterier i utøvende musikk er et tjenlig verktøy for å fremme pålitelig vurdering av studentenes bachelorkonsert, har jeg stilt følgende tre spørsmål: Hvilke forventninger og forståelser legger sensorer til grunn i møte med studentenes bachelorkonsert? Hvordan erfarer sensorer forhåndsdefinerte kriterier i møte med studentenes bachelorkonsert? I hvilken grad har sensorene nytte av karakterbeskrivelsene?

Sentralt i denne studien står begrepsparet holistisk og analytisk (kriteriebasert) vurdering. Med en holistisk modell utgår sensorer fra helhetsinntrykket av studentens framføring og setter en karakter. Med en analytisk modell utgår de fra forhåndsdefinerte kriterier. I det følgende vil jeg sette eksamenskonserten inn i en utdanningskontekst.

Vurdering i utøvende utdanninger foregår primært i tre ulike situasjoner:

- (i) Som *diagnostiserende vurdering*, under inntaksprøver, der lærerpanelet lytter etter potensial og utviklingsmuligheter.<sup>2</sup>
- (ii) Som *formativ vurdering* (vurdering for læring), der tilbakemeldinger og individuell veiledning har som intensjon å styrke studentenes læring underveis.

1 Dette er et profesjonsråd under Universitets- og høskolerådet og er et samarbeidsorgan for de utøvende og skapende musikkutdanningene i Norge. Siden 2019: Nasjonalt fagorgan for utøvende og skapende musikk (FUM).

2 I studien til Sandberg-Jurström, Lindgren og Zanden i artikkelen «A Mozart concert or three simple chords? Limits for approval in admission tests to Swedish music teacher education» forekommer både diagnoser og prognoser over søkerens utviklingsmuligheter og potensial for å klare utdanningen.

- (iii) Som *summativ vurdering* (vurdering av læring), som er den formelle informasjonen til omverdenen om læringsutbyttetets kvalitet. Den gir informasjon om studentenes kompetanse, ofte på slutten av et opplæringsløp (Stobart & Gipps, 1997; Harlen, 2006). Her faller bachelorkonserten inn. Den har en form for eksternt kontrollsystem, med ekstern sensor for bedømming.

Det formaliserte vurderingssystemet for utøvende musikkutdanninger i Norge er vokst fram på bakgrunn av tradisjoner, kultur, særskilte vurderinger av kunstnerisk kvalitet og epistemologiske antagelser. Vurderingspraksiser må ses på bakgrunn av dette. Hovedinstrumentfaget er karakterisert av en individualisert undervisningsform, der den formative vurderingen er en integrert del av samhandlingsmønsteret mellom lærer og student. Å være lærer i faget innebærer å utvikle et personlig, nært og subjektivt forhold til studenten. Dette står i motsetning til bachelorkonserten, der det stilles krav om en ikke-subjektiv og distansert holdning. Opplæringstradisjonen kjennetegnes av instruksjon, ofte med søkelys på teknisk og uttrykksmessig artisteri. Den hviler på informert bedømming fra ekspertensorer som kommer til konsensus, der den eksterne sensoren ofte har hatt siste ord i tvilstilfeller. Sensorene er vanligvis selv utøvere på ekspertnivå.

Bachelorkonserten handler om å kvalitetsvurdere studentens beherskelse av et repertoar, i sanntid, med kun ett forsøk. Etter at studenten har jobbet med sin kunstneriske utvikling i fire år, gis det en formell vurdering basert på den ene prestasjonen. Det er ikke knyttet klagerett til den formelle vurderingen. For studenten blir denne ene eksamenskonserten av desto større betydning. Sadler sier om den summative vurderingsformen at «it is essentially passive and does not normally have immediate impact on learning, although it often influences decisions which may have profound educational and personal consequences for the student» (Sadler, 1989, s. 120). Jeg anser det derfor som maktpåliggende at kvalitetsvurderingen oppleves pålitelig og troverdig, nemlig at man kommer fram til en karakter som kan rettfærdiggjøres, med basis i et eksplisitt vurderingsgrunnlag. En kvalitetsvurdering som ikke har tydeliggjort et vurderingsgrunnlag, kan bli oppfattet som esoterisk viten, en type kunnskap som andre ikke har innsyn i.

Det er en rekke utfordringer knyttet til vurdering av kunstneriske prestasjoner generelt. Det er for eksempel et misforhold mellom den musikalske erfaringen og ordene vi bruker for å beskrive den (Johnson, 1997, s. 272). Musikkutøvelsen har en innebygd kompleksitet, der mange kvalitative og skjønnsmessige aspekter, som klangfarge, følsomhet, nerve, frasering, ekspressivitet og artisteri, gjør den krevende å samtale om. Disse refererer Sadler (1989, 2009b) til med begrepet «multiple criteria». I tillegg til de vanskelige verbaliserbare dimensjonene er det et komplisert samvirke mellom dem, noe som gjør at den overordnede utførelsen innebærer noe mer enn summen av delene. Noen av kriteriene er uskarpe, preget av gradvis overgang fra en tilstand til en annen, i motsetning til fag der epistemisk kunnskap vurderes, og der overgangene er diskontinuerte. De mange skjønnsmessige aspektene ved musikkutøving er eksempler på abstrakte konstruksjoner med tvetydig mening som må ses avhengig av konteksten. Svaret på kvalitet ligger heller ikke i hver sin ende av et kontinuum av riktig eller galt. I studentens utøving formidles en intensjon: På den ene siden er idealet et personlig uttrykk, på den andre siden skal det være innenfor stilistiske rammer.

Sluttvurderingen gis i form av en bokstavkarakter (klassisk og kirkemusikk) eller som «bestått» eller «ikke bestått» (jazz). Karakteren skal reflektere nivået som er nådd, med basis i standarden for bachelorexamen, uavhengig av innsats eller læringsprosess. Sadler (1989, s. 129) definerer en standard slik: «A standard or reference level is a designated degree of performance or excellence.» Kvalitetsnivået vil forventes å være ulikt ved en bachelorexamen og en mastereksamen. Karakteren ledsages gjerne av en samtale der den formelle vurderingen legitimeres. Slik kan den summative vurderingen ha formative elementer i seg (Harlen, 2006). Følgende vurderingskriterier legges til grunn for karakterfastsettelse og tilbakemelding til studenten i utøvende musikk:

## Bachelor, klassisk og kirkemusikk

- musikalsk bevissthet (kunstnerisk uttrykk, tilstedeværelse, karakter, stilforståelse, tolkning)
- kommunikasjon (samspill, publikumskontakt, presentasjon og formidling)

- instrumentbeherskelse (klangbehandling, frasering, dynamikk, rytmikk, teknikk)
- konserten som helhet (dramaturgi, repertoar, konsentrasjon)

## Bachelor, jazz

- musikalsk bevissthet (gehør- og rytmikkvalitet, harmonisk forståelse, stilistiske aspekter, kreative idéer, formbevissthet, dynamisk sensitivitet)
- sanntidskvaliteter (improvisasjonsevne, nærvær)
- samspillferdighet (rollebevissthet i ensemblet, kommunikasjon, bandleidelse)
- instrumentkontroll (klangbehandling, teknikk, variasjon)
- konserten som helhet (uttrykksvilje, konsentrasjon, sjangeroverskridelse, kunstnerisk og konseptuell ambisjon) (Rådet for utøvende musikk, 2018)

Karakterbeskrivelsene har implementert vurderingskriteriene som forventet læringsutbytte. De er uttrykt som ulike kvalitetsnivåer ved prestasjonen, fra så svakt til stede at man ikke anser læringsmålene som nådd (F), til en fremragende prestasjon (A) (se vedlegg). De samme prinsippene ligger til grunn for «bestått» eller «ikke bestått».

I det følgende presenterer jeg tidligere forskning og teoretisk rammeverk for artikkelen. Deretter kommer fokusgruppeintervjuet med datainnsamling, resultater og analyse, med noen mulige implikasjoner av studien helt til slutt.

## Tidligere forskning

Forskningslitteraturen viser ulike oppfatninger av formell vurdering i utøvende musikk.

Det har blant annet vært en viss motvilje i musikk- og bildekunstmiljøer (Olsson, 2010; Sandberg, 1996), da karaktersystemer har vært ansett som reduksjonistiske på grunn av de estetiske fagfeltenes karakter. Teori rundt musikalsk utøving understreker at den er kompleks og av

interaktiv natur, der mange faktorer påvirker hverandre samtidig som den skjer i sanntid (Sadler, 1989; Saunders & Holahan, 1997; Swanwick, 1998). Vurdering basert på helhetsinntrykk, ofte personlig fundert og med få begrunnelser, har derfor karakterisert utøvende instrumentalmiljøer. Ifølge Soep (2004) vil en eksamenskarakter kunne gjenspeile bare deler av helheten, og det vil bli forenkende forklaringer. Andre (Johnson, 1997) har vist til et semantisk problem ved å oversette det musikalske språket til verbalspråk, som aldri kan fange opp nyansene og hvordan parametrene virker sammen. Dette til tross for en rekke studier som understøtter innføringen av kriteriebasert vurdering i utøvende musikk.

Sadler (1985, 1989, 2005, 2009a, 2009b, 2015) har vært sentral i den vurderingsteoretiske diskursen, med klargjøring av vurdering som konsept, funksjonen av vurderingskriterier og med et kompatibelt syn på begrepsparet kriteriebasert og holistisk. Saunders og Holahans (1997, s. 270) studie av 926 søkere til Connecticut All-State Band, som ble vurdert av 36 instrumentaldommere, viser at kriteriespesifikk vurdering kan brukes i vurdering av treblåser- og brasstudenters framføringer, med substansiell reliabilitet. Vurderingsgrunnlaget var tonekvalitet, teknikk og artikulasjon, rytmisk nøyaktighet, interpretasjon og bladlesing. Studien viser til at det framkommer mindre informasjon når alternative vurderingsformer er brukt. Gynnild (2010) gjør en kasusstudie som retter oppmerksomhet mot bruken av kriterier i vurdering av sang, der faglærerne mener redskapet bidrar til en mer konsistent vurdering. Flere har funnet høye nivåer av reliabilitet ved bruk av kriteriespesifikk vurdering (Azzara, 1993; Bergee, 2003; Rothlisberger, 1993). En intervjustudie med 15 lærere ved Sydney konservatorium (Stanley et al., 2002) kartla hvordan lærerne beskrev holistiske og kriteriespesifikke tilnæringer til vurdering. Funnene indikerer at kriterier gir et nyttig fokus, idet de bidrar til å artikulere ønskelige framføringsaspekter som tilbakemeldinger til studentene. Studien konkluderer også med å understreke behovet for at sensorer trenes i et komplementært syn på holistisk og kriteriebasert vurdering, og særskilt i prosedyrer for bruk av kriterier der nivået på studentens framføring varierer i løpet av en solokonsert. Også i Newsome (2015) sin studie, som ble lagt til grunn for utarbeidelsen av de nasjonale sensorveiledningene, argumenteres det for at instrumentalutøvende

miljøer bør finne en vurderingsmodell som balanserer «the traditional» og «the academic model», med særskilt henvisning til at det vil kunne hjelpe studentene i øvingsarbeidet. Samtidig viser hun til at ren kriteriebasert vurdering kan gjøre at man går glipp av viktige estetiske aspekter som det er umulig å verbalisere. Hun anbefaler derfor et kriteriesett som først utfordrer sensorer til å ta stilling til konserten som helhet, før de tar stilling til et sett med vurderingskriterier.

Vinge (2011) argumenterer for å behandle relasjonen mellom kriteriebasert og holistisk vurdering som et kontinuum istedenfor to adskilte modeller. Ser man til studier fra andre fag som har kommet langt innen vurderingsforskning, for eksempel tekstvurdering, har det vist seg at en «nedenfra og opp»-tilnærming med standardiserte vurderingskategorier gir større grad av pålitelighet og samsvar i vurderinger enn «ovenfra og ned»-tilnærminger (Matre & Cameron 2018). I dag er høyere musikkutdanninger preget av store forventninger om forskning, publikasjoner, standardisering av emner, grader og posisjoner (Angelo et al., 2019). Hensynet til studentens rettssikkerhet er fremmet med større styrke.

## Teori

Mitt teoretiske utgangspunkt er Sadler og hans begrepspar holistisk og analytisk (kriteriebasert) vurdering. På mange måter argumenterer Sadler for begge modellene, alt etter hvor han er i sin vitenskapelige produksjon. Med den holistiske modellen setter sensorer en karakter basert på helhetsinntrykket av studentens framføring, for så å gi den substans ved å referere til detaljer (Sadler, 2005, s. 179). Vanligvis landes et karakterforslag raskt (Sadler, 2009b, s. 46). Sadler karakteriserer vurderingsmåten som intuitiv eller impresjonistisk. Holistisk vurdering får gjerne karakter av private og subjektive refleksjoner: «With a global approach judges usually employ personally selected criteria to classify, score, or rank a performance in relation to others» (Stanley et al, 2002, s. 47). Thompson og Williamon (2003, s. 26), peker på at erfarne musikere utvikler et slags indre mentalt system, basert på svært høy ekspertise, og til og med et som er spesifikt for det stykket som blir framført. Sadler (2009b) viser til en mulig begrensning ved den holistiske modellen som følge av at sensorer

kan ha ulike standarder, og av at kjennskap til enkeltstudenters personlighet kan blande seg inn eller føre inn fordommer.

Den kriteriebaserte modellen tar utgangspunkt i en prosedyre der sensorer gjør vurderingen med basis i forhåndsdefinerte kriterier. Disse ses i lys av standarden vurderingshandlingen foregår ut fra. Sadler (1989) definerer kriterier som dimensjoner ved framføringen som er relevante for vurdering. I hvert kriterium ligger et spekter av ulike kvalifikasjonsnivåer. Om man antar at instrumentkontroll og samspillferdighet er sentrale kriterier ved både bachelor- og masterkonserter, vil forventningene måtte tilpasses to ulike, standardiserte nivåer. Sadler (1989) beskriver hvordan kriterier tjener flere viktige funksjoner. For det første minimaliserer de tilstedeværelsen av vilkårlighet og bidrar til detaljerte og eksplisitte tilbakemeldinger til studentene gjennom systematikk. De er nyttige for klargjøring og kommunikasjon, gir substans og konsentrert oppmerksomhet. Videre forenkler kriteriene bedømmelsesprosessen idet de reiser en vurderingsstrategi som lar seg overføre på flere i samme utdanningsklasse. De kan fungere som et verktøy for å undersøke eventuell uenighet i vurderingsprosessen, og et hierarki av utdypende vurderingskriterier viser hvor det kan være aktuelt med kompromisser.

Innvendingene mot forhåndsoppsatte kriterier går på at de ikke klarer å reflektere alle aspekter ved en framføring, og at de kan bli en overforenkling av det sensor hører. Vinge (2011, s. 206) er inne på det samme når han påpeker at det ofte vil være umulig å fastsette klare læringsmål eller vurderingskriterier i musikkfaget, av den enkle grunn at man ikke kan forutse hva elevene «bringer til bordet». Studenters bachelorkonsert kan ha usammenlignbart repertoar, og kriteriene kan falle ulikt ut for ulike sjangre og ulike instrumentgrupper. Kriterier som stilforståelse, uttrykksvilje og intonasjon kan substansielt fortone seg ulikt fra stykke til stykke. Noen kriterier vil være relevante for visse instrumenter og mindre relevante for andre. Intonasjon på et høytpresenterende nivå må vurderes ulikt i for eksempel Bernsteins «I hate Music», i Mozarts «Alleluia» og i en blues. «Professional qualitative judgement consists in knowing the rules for using (or occasionally breaking) the rules» (Sadler, 1989, s. 124).



## Kvalitativ vurdering av komplekst læringsutbytte i musikk

Kvalitativ vurdering i fag med komplekst læringsutbytte er forankret i aksiologiske verdier (Sadler, 1985, s. 287). Disse er en type underliggende eller overbyggende verdier som er mer eller mindre ubevisste for oss selv, noe som stiller sensorer overfor betydelige utfordringer. Eksempler er sannhet, skjønnhet og rettferdighet, og i en bredere utdanningskontekst vil individuelt uttrykk og frihet være aksiologiske verdier. Verdiene er mye mer enn preferanser, smak og behag – snarere er de personlige overbevisninger. Som abstrakte ideer har de utviklet seg i et komplekst samspill av erfaringer, tradisjoner, verdsettelse og sosial diskurs. Disse verdiene kan forklare hvorfor vi betrakter visse kriterier som mer tungtveiende enn andre i vurderingshandlingen, eller hvorfor sensorer fort kan lande på en helhetsoppfatning av konsertframføringen, fordi kriteriene appellerer til en dypt forankret oppfatning av estetisk sensibilitet. Derfor er det viktig at forbindelsen mellom klasseromerfaringer og fundamentale menneskelige verdier drøftes og synliggjøres.

Grunnlaget for holistisk vurdering er innforstått kunnskap, som er basert på årelang erfaring. En musiker kjenner sitt instrument, har hørt en rekke versjoner av den samme trompetkonserten, kjenner den gjennom innstudering og utallige framføringer. Slike lærere prosesserer intuitivt det de hører, til komplekse avgjørelser og indikerer en karakter. I denne studien er «konserten som helhet» definert som et eget vurderingskriterium. Til syvende og sist er «kvalitet» et integrert konsept som består av uendelig mange ulike kriterier, og som karakteriserer et arbeid som helhet (Sadler, 2009b). Sentralt i vurderingen av kvalitet står gjenkjenning. Sadler nærmer seg her det Polanyi (2000) betegner som «tacit knowledge», altså at mennesker sitter inne med en betydelig mengde kunnskap som ikke lar seg artikulere fullt ut. Polanyis legendariske eksempel er gjenkjenning av et ansikt i en stor folkemengde. Komplexiteten i ansiktet gjør det vanskelig å sette ord på hva vi kjenner igjen, men ikke vanskeligere enn at dagens fantomtegnere blir stadig dyktigere til å konstruere ansikter som ligner mer og mer en verbal beskrivelse. Tilsvarende kan utøvende musikkmiljøers sensorer styrke verbaliseringskompetansen ved å la kriterier spille en

viktig rolle i kvalitativ vurdering. Resultatet vil bli deres «fantomtegning» av den utøvende prestasjonen som blir vurdert.

## Metode

Datagenereringen har et kvalitativt design basert på fokusgruppeintervju. Fokusgruppeintervjuer framholdes som velegnet når man ønsker å utvikle kunnskap som gir en dypere forståelse av meningsdannelse i gruppen (Madriz, 2000; Wibeck, 2010). Min hensikt med dette var å belyse vurderingsfeltet gjennom en fri uttrykksform der deltagerne kunne dele erfaringer, utveksle meninger og historier relatert til vurderingsutfordringene de alle kjenner til. På den måten ville jeg legge til rette for dataskaping i fellesskap. Samtaleformen bidrar til spontane svar, slik at selve interaksjonen kan generere data.

Utvalg for studien er seks sensorer som var engasjert til bacheloreksamen våren 2019. Det at de var tilgjengelige til samme tid, gjorde det mulig å samle dem til samtale rundt vurderingstematikk. De ble kontaktet en måned i forveien, med invitasjon til å delta. De ble informert om hensikt og metode, om at det ville bli gjort opptak, og om at materialet ville bli aidentifisert og konfidensielt behandlet. Prosjektet var meldt til NSD – Norsk senter for forskningsdata. Alle ønsket å delta. Sensorene representerer en homogen fagkultur som utøvere på høyt nivå: fem med kunstmusikkbakgrunn og en med jazzbakgrunn. Likevel utgjør de et spekter hva angår instrumentfordeling, sjangerfordeling og erfaring. De representerer instrumentene klassisk sang, jazzsang, trombone og orgel. En oppga å være sensor for første gang, to hadde lang erfaring som intern sensor, mens tre oppga årelang erfaring som både intern og ekstern sensor i og utenom norske utdanningsinstitusjoner. Fire var studentenes faglærere. Den formelle sensorveiledningen de brukte under eksamen, er utviklet med tanke på kunstneriske fellesnevner for et standardisert bachelornivå (Rådet for utøvende musikk, 2018).

Intervjuet ble gjennomført dagen etter at de hadde sensurert studentenes bacheloreksamener. Instituttets studio ble stilt til disposisjon, og en lydtekniker gjorde opptaket. En førstekonsulent fra instituttets eksamensadministrasjon deltok i tilfelle det skulle oppstå spørsmål rundt

selve eksamensavviklingen. Jeg fungerte som moderator, ønsket velkommen og ledet intervjuet. To momenter ble presisert innledningsvis: At alle skulle forsøke å gi hverandre plass i samtalen, og at det her ikke finnes «rette eller gale» svar. Dette mener jeg forhindret asymmetri mellom deltagerne, og et trygt klima ble etablert. Jeg betrakter det også som et godt utgangspunkt at jeg har en tilsvarende musikkutdanning som intervjuobjektene, og at jeg selv er lærer. Å være på innsiden av tematikken skaper tillit og gjør det lett å følge opp tankegods. Etter intervjuet tok jeg og administrasjonens representant et tilbakeblikk på intervjuet for å oppsummere vårt helhetsinntrykk. Fokusgruppeintervjuet ble deretter transkribert i detalj. Med en liten gruppe sensorer har jeg forsøkt å ivareta forskningsetiske hensyn ved å unngå opplysninger som kan identifisere kildene. Av etiske grunner er også alle sitater oversatt til norsk. Jeg har tillatt meg å stå fritt med tanke på hvilket kjønn jeg tilskriver mine informanter.

Analysen av datamaterialet ble gjort ut fra Giorgis fenomenologiske analysemetode slik den er beskrevet i De Castro (2003), supplert med et hermeneutisk perspektiv (Gadamer, 2010). Metoden åpner for å undersøke både opplevelser, erfaringer og meninger hos sensorene for å beskrive fenomenets essens: vurderinger av kunstnerisk kvalitet. Sensorenes forståelser vokser samtidig fram i relasjonen mellom dem underveis i intervjusamtalen. Tanker om vurderingsprosessen blir dermed noe som også forhandles fram gjennom dialog idet de pendler mellom førforståelse og forståelse, helhet og del. I beste fall oppnår man at gruppas «tacit knowledge» blir gjort eksplisitt om faktorer som kan virke konstituerende på vurderingspraksis. Før min egen utforskning av materialet var det et poeng å ta et skritt tilbake for å tilstrebe undring og åpenhet og å sette min egen førforståelse eller subjektivitet i parentes. I første trinn av analysearbeidet ble det transkriberte intervjuet lest med tanke på å gripe helheten og fange opp tendenser av meninger og mønstre i samtalen, i tråd med fenomenologisk reduksjon: Hva er det som trer fram? Flere historier om bruken av karakterskalaen dukket opp, samt et tilbakevendende ønske om at verktøyene må brukes til å gi støttende tilbakemeldinger til studentene. Så ble intervjuet systematisk gjennomgått for å identifisere meningsbærende elementer som ser ut til å dominere i gruppen, og som

svarer på forskningsspørsmålene. Dette resulterte i koding av sentrale utsagn som var relevante for fenomenet som skulle utforskes. «Å zoome ut og zoome inn» var et slikt gjennomgående utsagn. Uttalelsene ble analysert og transformert ved å gjengi deltagerens realitetsoppfatning i et fagspråk for å trenge inn i de dimensjonene og utfordringene man møter når man skal verbalisere kvaliteten i det kunstneriske uttrykket.

## Resultater

Resultatene fra intervjuet har gitt innsikt i følgende tre analysekategorier:

- (1) forventninger til eksamenskonserten
- (2) kriteriebasert eller holistisk vurdering
- (3) karakterbeskrivelsenes bidrag i vurderingsarbeidet

Teksten videre er strukturert med disse kategoriene som underoverskrifter. Sensorene benevnes med S1, S2 og så videre. For å være tro mot sensorenes forståelsesprosess vil jeg i det følgende rette søkelyset mot sitater fra intervjuet.

### Forventninger til eksamenskonserten

Fire av sensorene sier de ikke stiller med forventninger til eksamenskonserten, eksemplifisert av utsagnet nedenfor. To kommenterer ikke dette, men er avventende.

S1: Jeg forventer ingenting. Jeg prøver å se hva studenten tilbyr ... utgår fra det. Så sammenligner jeg forventninger etterpå, jeg prøver å ikke sette opp noe på forhånd. Noen har en konsert rettet mot moderne musikk, ensemblestoff. Noen holder seg innen visse stiler. Noen kan kommunisere med barn. Jeg prøver å la liksom alt komme inn, så jeg prøver å være ganske nullstilt.

Utsagn av denne art viser en av utfordringene ved et sett definerte kriterier som skal gjelde for alle, uavhengig av instrument og repertoar. Samme sensor bemerker også viktigheten av samtalen med intern sensor etter konserten: «Da spør jeg absolutt hvordan det her tilsvarer i hans

øyne, hvordan den her prestasjonen tilsvarer det forventede. Hevdet det seg? Sank det? Eller var det som forventet? Jeg synes det må også være menneskelig.»

S2 er opptatt av overgangen fra det å ha vært studentens lærer i tre år, til å lytte uten forventninger på denne ene eksamensdagen: «Vi faglærere må nullstille oss. Dette er det som gjelder, å lytte og se med blanke ark. Dette har vi fått streng beskjed om, så det du har tenkt tidligere, kan du bare glemme». S4 poengterer at han er sensor for første gang og dermed ikke har eksamenskonteksten å relatere seg til og følgelig ingen konkrete forventninger. Men som aktiv musiker i en årrekke mener han at utøvererfaringen likevel gjør ham forberedt:

S4: Jeg kan bare snakke om det jeg opplevde i går, jeg snakker om én erfaring. Jeg ante ikke hva jeg kunne forvente ... eller ... jeg hadde jo en forventning om å få en opplevelse. Men på hvilket nivå, hadde jeg ingen anelse om. Selv om man ikke lager bevisste forventninger, så har man noe i hodet, en slags uartikulert forventning. Og om man ikke har vært aktiv pedagog, så har man vært aktiv sanger i mange år og hører vel forskjellen på sunt og usunt.

«Sunt og usunt» relaterer etter alt å dømme til studentens stemmebruk og peker mot eksperters intuitive måter å prosessere et auditivt inntrykk over i en karakter på, uten å gå veien om vurderingskriterier.

## Kriteriebasert eller holistisk vurdering

Samtalen indikerer en samstemthet blant sensorene når det gjelder å ønske forhåndsdefinerte kriterier velkommen. Flere mener at de under eksamen har et lyttfokus som kombinerer helhet og enkeltparametre, og at det foregår en vekselvirkning mellom dem. Likevel kan enkelte uttalelser antyde at de starter med helhetslytting først, for deretter å spesifisere hva denne består av. Følgende illustrer dette:

S1: Ser vi på det som helhetssyn eller som detaljert? Der er egentlig samme ting, det er bare et spørsmål om hvor mye du zoomer inn og zoomer ut. Zoomer du ut, så hører du helhetsinntrykket. Og når du skal gjøre vurderingen, så er du nødt til å zoome inn. Hva var det for mangler som gjorde at den kunstneriske tingen der ikke fungerte? Hva var det som gjorde at det var veldig godt? Da får

man zoomer inn og si at de kvalitetene var helt fantastiske, men de manglene fantes. Så altså: Selvfølgelig synes jeg at det er veldig godt at dette er på papir. Men, altså, både i eksamenssituasjonen og når vi underviser våre egne studenter, så er det jo de her kriteriene vi driver med hele tiden. Den perfekte helheten består jo av en masse perfekte detaljer. Nei, det fins ingen konflikt her, synes jeg, i det hele tatt.

Uttrykket «å zoomer inn og zoomer ut» er tilbakevendende i samtalen. S4 fastslår at kriteriene hadde hjulpet veldig med å klargjøre egne tanker før konserten, særskilt på bakgrunn av at han var engasjert som sensor for første gang. S3 (representerer utøvende jazztradisjon) bringer til torgs betydningen av improvisasjon. Dette er trolig et særskilt viktig vurderingskriterium: «Det vil alltid være et fokus på improvisasjonsevnen, uansett hvilken type musikk jazzstudentene velger å framføre på eksamen, så er det dette å ha det helhetlige overblikket, det ligger der alltid.»

## Karakterbeskrivelsenes bidrag i vurderingsarbeidet

Sensorene problematiserer det de kaller en «inflasjon i toppkarakterer» i utøvende utdanninger, og at studentene faktisk blir skuffet når de får en C. Etter at de har lest karakterbeskrivelsene før eksamen, gir to av dem uttrykk for at det var oppklarende å se at gjennomsnittskarakteren C indikerer en prestasjon som viser god kunstnerisk kapasitet og ingen åpenbare mangler. Det vil gjøre det lettere for dem å bruke C i større grad, og å legitimere den. S1 påpeker at han hadde lest opp for en kandidat hva som sto under en D, noe som bidro til at kandidaten forsto at C var et bra resultat. Dette følges opp av S6: «Veldig ofte kommer vi ut med en A, A, B, B, B, A og har ikke sett skyggen av en C. Jeg var tilfreds i går da vi hadde både en B og en C.»

S3 bringer inn en erfaring med en sensor som ville stryke en student fordi hun ikke hørte noe improvisasjon, men opplevde musikken som komponert. Dette i motsetning til den interne faglæreren som hørte mange improviserte elementer. Det hadde ført til en lengre diskusjon. Her ligger flere dilemmaer: Hvor går skillet mellom improvisert og gjennomkomponert musikk? Hvor mye skal det improvisatoriske elementet vektlegges når studenten ellers viser tekniske og uttrykksmessige ferdigheter

på et habilt nivå? Hvor går grensen mellom «bestått» og «ikke bestått» i en vurderingsskala med kun to karakterer?

Samtalen ledet etter hvert over på hva studentene kommer til å leve av etter endt studium, om svaret skal ha konsekvenser for hva man vurderer, og om urimeligheten ved å vurdere kun én sluttprestasjon i lys av tre–fire års arbeidsinnsats. Her var oppfatningene differensiert. Et ytterpunkt representeres av S4, som framhever prøvesang for profesjonelle musikere, og «det du presterer der og da, er det som er det avgjørende», noe som legitimerer en tilsvarende eksamen. Det motsatte synspunktet hevdtes av S1:

S1: Jeg er enig, sånn ser musikerlivet ut, jeg har selv vært orkestermusiker i 40 år. Ting er sånn her: Det kommer jobb ledig i orkesteret. Det kommer 60–70 søkere. Det er én vinner og 69 tapere. Sånn må det ikke være i en utdannings-situasjon. Går man til et orkesterspill eller livekonsert, da er det der og da det gjelder. Hvor mange av bachelorstudentene som går ut hvert år i Norge, kommer til å leve som utøvende instrumentalister? Dette er en skole, det er ikke en solistkonkurranse. Det er ikke «livet er hardt, og det skal vi vise».

S5 verdsetter de musikktilpassede karakterbeskrivelsene, men uttrykker misnøye med standpunkt-karakterer som forsvant. Disse avspeilte studentens innsats over tid. Flere er enige, og følgende uttalelse er representativ:

S1: Hva skjedde med den kontinuerlige evalueringen? Er fire års arbeid intet verd hvis man har en dårlig dag? Forhåpentligvis kan sensorer være så smarte og så erfarne at man hører at her er et menneske som faktisk underpresterer i dag. Da må man telle det til kandidatens fordel. For jeg synes det er helt umenneskelig ... om det er fire års arbeid, så er det plutselig en voldsom deep, og så skal den henge igjen resten av livet.

Både S1, S4, S5 og S6 tar til orde for at antall sensorer bør være flere enn to. De bruker ord som rettssikkerhet og rettferdighet ved å være tre. Flere foreslår en gjennomgående sensor på alle eksamener:

S1: Vi snakker om rettssikkerhet på studentens vegne ... vi var fire messing-blåsere i går. Om en middels god student skal ta sin eksamen, og det er to andre genier, da er det veldig viktig at man sammenligner mot disse kriteriene og ikke

bare i forhold til hverandre. Om man har en gjennomgående sensor på huset som kan sikre en så objektiv bedømming til karakterskalaen som mulig, så tror jeg det er ønskelig.

S6 påpeker at «for en ekstern sensor er det fare for at en kan føle seg presset av den læreren som har hatt studenten», og da er det fint å være tre. S5 rapporterer gode erfaringer med å trekke studentene selv inn i vurderingsarbeidet gjennom studieåret: «Mange studenter har et ganske sunt syn på hvor de står, og kan godt bli med i den debatten. Dette skjer jo faktisk, i realiteten, hver eneste time.»

## Diskusjon

Sensorgruppas samtale rundt sine «forventninger» til bachelorkonserten avdekker ulike forståelser av hva som ligger i begrepet forventninger, og det kan diskuteres om spørsmålet de fikk, var presist nok. I et hermeneutisk perspektiv møter nemlig ikke sensorene forutsetningsløse til konserten. De har for eksempel rike utøvende erfaringer, de har lest og anvendt kriteriebeskrivelsene, og de vil veksle mellom sin førforståelse av oppdraget og den framforhandlede forståelsen i samtalene med faglærere etter konserten. S4 sin uttalelse er treffende: Fra å fastslå at han ikke møter med forventninger fordi dette var det første sensoroppdraget, korrigerer han seg selv i retning av å ha forventninger om en opplevelse, til å lande på at han faktisk har en uartikulert forventning. S2, som er faglærer, ser det som et mål å lytte uten forventninger, som en strategi for ikke-subjektivitet overfor studenter hun har fulgt i tre år. S1 representerer en tredje forståelse, nemlig ikke å lande noen forventninger før han har hørt studentens repertoarvalg og stilvalg, som kan ta ulike retninger. Sistnevnte har i en årrekke vært sensor ved bacheloreksamener, og vedkommende har sannsynligvis en taus kunnskap om standarden i ryggmargen, noe jeg tolker som en forklaring på hvorfor ikke bachelorstandard blir en videre tematikk på dette tidspunktet. Vedkommende vektlegger også samtalen med faglærer etter konserten om hvordan prestasjonene samsvarer med faglærers oppfatninger. Sadler (1985) reiser i en slik sammenheng spørsmål ved om altfor blanke ark kan bidra til at man ikke reflekterer



i stor nok grad over standarden for vurdering. Det indikeres for eksempel av utsagn som relaterer bachelorkonserten til framtidige prøvespill og prøvesang. Hva vurderes studentene i forhold til? Vurderes de ut fra om de skal bli profesjonelle musikere, eller vurderes de til et bredt yrkesliv? Det er nemlig grunn til å drøfte hvorvidt det legges et for smalt læringsutbytte til grunn for eksamen, så lenge det kun er bachelorkonserten som gjøres til gjenstand for formell vurdering med ekstern sensor. Biggs (1999) viser til at det er mer sannsynlig at studentene oppnår forståelse av fagstoffet når det er samsvar mellom læringsutbytte, læringsaktiviteter og vurderingsformer.

Sensorpanelet skal gjøre sin vurdering ut fra et på forhånd gitt sett med kriterier, der konserten som helhet er et eget kriterium. De uttrykker klart at de ikke ser noen motsetning mellom helhet og del. Sadler (1985) sier om dette at det sentrale er at man ikke betrakter kriteriene som separate, uavhengige dimensjoner, men som elementer av hensiktsmessighet. De må ses i en kontekst. Sensorene mener også at de kontinuerlig veksler mellom å lytte helhetlig og å lytte til de ulike elementene i framføringen. Uttrykket «å zoome inn og zoome ut» dekker dette. Kvale og Brinkmann (2009) viser hvordan en hermeneutisk meningsfortolkning nettopp fordrer en kontinuerlig bevegelse mellom del og helhet og mellom helhet og del, hvor førforståelsen hele tiden må settes på prøve og endres. En forsamtale sensorene imellom, hvor man reflekterer over standarden for bachelor, kunne inngå – og i et bredere yrkesperspektiv enn prøvespill.

Kriterier oppleves som nyttige idet de bidrar til å artikulere tydelig begrunnede tilbakemeldinger til studentene. Dette vil styrke påliteligheten og er i tråd med intervjuundersøkelsen til Stanley et al. (2002). Samtidig viser Sadler (2009b) til at man med ren kriteriebasert vurdering kan gå glipp av samvirket mellom estetiske aspekter, med begrunnelse i de uendelig mange kriteriene som til dels er tvetydige. Skjønt han fastslår at også den holistiske vurderingen i ettertid nødvendigvis må begrunnes med referanse til kriterier (Sadler, 2005, s. 179). Studentene må tidlig i studiet bli gjort kjent med vurderingskriteriene, diskutere dem og utvikle egen vurderingskompetanse. På det viset kan kriterier fungere som en felles referanse for øving, undervisning og

veiledning. Studentene trenger å høre et stort antall forbilledlige framføringer. Her kan den holistiske vurderingen være et utgangspunkt for å trene på operasjonalisering av begrepene artisteri og kunstnerisk kvalitet. På sin side forutsetter gode øvingsstrategier en kriteriebasert tilnærming.

Det er et spørsmål om ikke det å gjøre sensorers ekspertise til en form for «taus kunnskap» i lengden kan komme til å vanskeliggjøre analyser og saklige diskusjoner omkring vurdering. Rostvall og West (2003) advarer mot historisk legitimerede praksiser i instrumentalopplæringen, med en-til-en-undervisning, som en slags «black box», en privatisert aktivitet der kunnskap og metoder sjelden deles systematisk. Dette vil i verste fall kunne gjøre lærergruppen isolert og gi dem få muligheter for utvikling av undervisningsferdigheter. Profesjonsutøvelse krever også at man kan formidle og begrunne den. Eik (2013, s. 126) trekker fram språkliggjøring som en forutsetning for profesjonsutøvelse, og hun peker på at utøveren også må kunne sette ord på den delen av utøvelsen som preges av skjønn og improvisasjon. Mot denne bakgrunnen vil det å bygge en samtalekultur rundt kvalitet i konsertframføring kunne styrke sensorers og læreres vurderingspraksis og profesjonalitet.

Karakterbeskrivelsene synes å ha positive implikasjoner når det gjelder ettersamtalen med studentene. En indikator er sensorenes erfaring med å bruke karakterbeskrivelsene aktivt overfor studenten for å legitimere en karakter og en større del av skalaen. De fleste sensorene rapporterer erfart misnøye blant studenter som har fått karakteren C. Er det rett og slett kun en ukultur som har utviklet seg rundt eksamenskonsserter, at alt under A eller B forteller deg at du ikke er en habil musiker? Har studentene kunnskap om vurderingskriteriene og nivåbeskrivelsene i karakter-systemet? Om dette sier Gynnild (2010, s. 27):

Faktisk er det slik at de aller fleste tilfeller der det forekommer misnøye, er det slik at karakterkriteriene i liten grad er blitt kommunisert på forhånd, og at vurderingen derfor kommer som en overraskelse på studentene.

For å styrke rettsikkerheten til studentene må alle sider ved vurderingen gjøres så transparent som mulig. Studenten må ha innsikt i hvilke kriterier og premisser som legges til grunn, og han eller hun må være

sikker på at beslutningene skjer på en rettferdig og likeverdig måte. Å gjøre vurderingskriteriene kjent for studentene tidlig vil kunne styrke læringsutbyttet og kvaliteten i utdanningen. Sensorgruppa bruker selv begrepet rettsikkerhet i forbindelse med karaktersetting. Antall sensorer *har* betydning. Dette støttes av tidligere forskning (Bergee, 2003; Fiske, 1975, 1977). Bergees studie (2003) viser at påliteligheten øker med at vurderingspanelet øker, og gjerne til flere enn tre, aller helst til 5–7. Han tar også til orde for at ikke alle sensorene trenger å være spesialister på samme instrument som studentene. Sensorgruppa anser sluttsamtalen med studentene som det aller viktigste aspektet ved sluttvurderingen.

Dohn (2006) peker på det etisk og kompetansemessig problematiske i å være «intuitiv ekspert» i å sette karakterer. Det leder oss til spørsmålet om det trengs økt bevissthet rundt bruk av skalaen «bestått» eller «ikke bestått». I en rapport fra en arbeidsgruppe oppnevnt av Universitets- og høyskolerådet (2015) slås det fast at også denne skalaen skal relateres til i hvilken grad læringsutbyttet er oppnådd. I rommet mellom de to karakterene ligger en rekke nyanser av kvalitetsnivåer, men hvor ligger minimumsprestasjonen for å få «bestått»? Hva kjennetegner kvalitetsnivået til en student som ikke består? Kritiske samtaler, erfaringsdeling og tolkningsfellesskap vil kunne styrke bevisstgjøring rundt anvendelse av skalaen.

## Konklusjon

Resultatet av denne studien indikerer sensorgruppens støtte til kriteriebasert vurdering av studentenes bachelorkonsert. Kriterier gir potensial for en transparent, argumenterende og pålitelig vurderingspraksis. De har vist seg å være gode verktøy for tydelige og begrunnede tilbakemeldinger til studentene, men de må tilpasses instrumentkontekst, være standardorientert og ikke anvendes mekanisk. Det empiriske materialet peker også mot at de fagspesifikke karakterbeskrivelsene har vært til nytte for sensorene, og at de særskilt har gjort det lettere å legitimere en nyansert bruk av karakterskalaen overfor studentene. Imidlertid har denne studien et lite utvalg, den er hentet fra ett fagmiljø, og den har

begrenset instrumentrepresentasjon. Den viser derfor at det er behov for å gjennomføre tilsvarende forskning et annet sted med andre informanter for å kunne si noe om hvorvidt materialet også har holdbarhet i en større kontekst.

Det er behov for å knytte sensorers intuitive og holistiske kvalitetsoppfatning til språklig artikulert viten. Det vil komme både sensorer, lærere og studenter til gode, framfor å gjøre ekspertkunnskap til en esoterisk viten som kan komme til å hindre yrkesgruppens utvikling av refleksjons- og vurderingsferdigheter. Et rammeverk for sluttvurderingen synes å kunne styrke fagmiljøene, der læringsutbytte, mål, vurderingsformer og vurderingsskala henger godt sammen og er i tråd med et bredt læringsutbytte og et allsidig yrkesliv.

Det gjenstår mye forskning på hvordan sensorer faktisk ender opp med å gi den karakteren de gir, og på hvordan de vekter ulike prestasjoner i løpet av en students eksamenskonsert. Andre sentrale spørsmål er i hvilken grad sensorer vurderer ved å sammenligne med andre studenter på samme instrument, eller ved orientering mot læringsutbyttebeskrivelser. Å arkivere sensorers vurderingsmateriale fra eksamenskonsserter kan derfor vise seg å bli et godt empirisk grunnlag for videre forskning i vurderingsfeltet.

Denne studien peker mot følgende mulige implikasjoner for praksis ved utøvende musikkutdanninger:

- Å revurdere små sensorgrupper av hensyn til rettferdighet og pålitelighet.
- Å vurdere et bredere, formelt vurderingsgrunnlag til bacheloreksamensamen.
- Å trekke studenten aktivt inn i vurderingspraksiser.
- Å utvikle et artikulert tolkningsfelleskap eller fagfellefelleskap for å styrke refleksjonene rundt hvordan kvalitet i utøvende musikk skal bedømmes, og hvordan man kan gjøre disse bedømmingene transparente.

Det vil kreve substansiell utstaking av en kurs der pedagogisk styrking og bevisstgjøring av det utøvende fagpersonalet er et mål.

## Referanser

- Angelo, E., Varkøy, Ø. & Georgii-Hemming, E. (2019). Notions of mandate, knowledge and research in Norwegian classical music performance studies. *Journal for Research in Arts and Sports Education*, 3(1), 78–100.
- Azzara, C. D. (1993). The effect of audiation-based improvisation techniques on the music achievement of elementary music students. *Journal of Research in Music Education*, 41, 328–342.
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137–150.
- Biggs, J. B. (1999). *Teaching for enhanced learning: What the student does*. Society for Research into Higher Education, 18, 57–75, Open University Press.
- De Castro, A. (2003). Introduction to Giorgi's existential phenomenological research method. *Psicología desde el Caribe*, 11, 45–56. <https://www.redalyc.org/pdf/213/21301104.pdf>
- Dohn, N. B. (2006). Karaktergivning – intuitiv ekspertise eller «viden i praksis»? *Dansk Universitetspædagogisk Tidsskrift*, 1(1), 38–46.
- Eik, L. T. (2013). *Ny i profesjonen. En observasjons- og intervjustudie av førskolelærere videre kvalifisering det første året i yrket* [Doktorgradsavhandling, Universitetet i Oslo]. <https://www.duo.uio.no/bitstream/handle/10852/44082/draavhandling-eik.pdf?sequence=1&isAllowed=y>
- Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education*, 23, 186–196.
- Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education*, 25, 256–263.
- Gadamer, H.-G. (2010). *Sannhet og metode: grunntrekk i en filosofisk hermeneutikk* (L. Holm-Hansen, (oversetter)). Pax Forlag.
- Gynnild, V. (2010). *Kriterier og skjønn i evaluering. En kasusstudie i utøvende musikkutdanning* (NOKUTS rapporter nr. 2010 – 4). [https://evalueringsportalen.no/evaluering/kriterier-og-skjonn-i-evaluering-en-kasusstudie-i-utovende-musikkutdanning-oktober-2010/Gynnild\\_Kriterier\\_og\\_skjonn\\_i\\_evaluering\\_en\\_kasusstudie\\_2010.pdf/@@inline](https://evalueringsportalen.no/evaluering/kriterier-og-skjonn-i-evaluering-en-kasusstudie-i-utovende-musikkutdanning-oktober-2010/Gynnild_Kriterier_og_skjonn_i_evaluering_en_kasusstudie_2010.pdf/@@inline)
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. I J. Gardner (Red.), *Assessment and learning* (s. 103–118). Sage Publications.
- Johnson, P. (1997). Performance as experience: The problem of assessment criteria. *British Journal of Music Education*, 14(3), 271–282.
- Kvale, S. & Brinkmann, S. (2009). *Interview – en introduktion til et håndværk* (2. utg.). Hans Reitzels Forlag.
- Madriz, E. (2000). Focus groups in feminist research. I N. K. Denzin & Y. S. Lincoln (Red.), *Handbook of qualitative research* (2. utg.), 835–850. Sage Publications.

- Matre, M. E. & Cameron, D. L. (2018). Fører opplæring i felles vurderingskriterier til større grad av enighet? En studie av tekstvurdering som en del av spesialpedagogutdanning. *Acta Didactica Norge*, 12(4), art. 11.
- Meld. St.1 (2016–2017). *Kultur for kvalitet i høyere utdanning*. Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/meld.-st.-16-20162017/id2536007/>
- Newsome, E. (2015). A search for balance. The development of a performance assessment form for classical instrumental music in the tertiary context. *Assessment in Music Education: From Policy to Practice*, 16, 153–170.
- Olsson, B. (2010). Bedømmning i estetiske ämnen – mer än bra eller dålig konst eller musik. I A. K. Ekholm (Red.), *Bedømmning för lärande – en grund för ökat kunnskande* (hefte 3, s. 44–54). Lärarförbundet, Stiftelsen SAF.
- Polanyi, M. (2000). *Den tause dimensjonen. En introduksjon til taus kunnskap*. Spartacus.
- Rostvall, A.-L. & West, T. (2003). Analysis of interaction and learning in instrumental teaching. *Music Education Research*, 5(3), 213–226.
- Rothlisberger, D. J. (1993). The effects of video tape preparation on performance anxiety and its relationship to instrumental music achievement among high school students. *Dissertation Abstracts International*, 53(7), 2287A.
- Rådet for utøvende musikk. (2018). *Rapport om sensorveiledninger i høyere musikkutdanning* (Rapport). <https://s3.eu-north-1.amazonaws.com/nmh-nettsted/files/Studier/Sensorveiledninger-i-h%C3%B8yere-musikkutdanning.pdf>
- Sadler, D. R. (1985). The origins and functions of evaluative criteria. *Educational Theory*, 35(3), 285–297.
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191–209.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175–194.
- Sadler, D. R. (2009a). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179.
- Sadler, D. R. (2009b). Transforming holistic assessment and grading into a vehicle for complex learning. I G. Joughin (Red.), *Assessment, learning and judgement in higher education*. [https://link.springer.com/chapter/10.1007/978-1-4020-8905-3\\_4](https://link.springer.com/chapter/10.1007/978-1-4020-8905-3_4)
- Sadler, D. R. (2015). Backwards assessment explanations: Implications for teaching and assessment practice. I D. Lebler, S. Harrison & G. Carey (Red.), *Assessment in music education: From policy to practice* (s. 9–19). Springer.
- Sandberg-Jurström, R., Lindgren, M. & Zandén, O. (2021). A Mozart concert or three simple chords? Limits for approval in admission tests for Swedish specialist music teacher education. I E. Angelo, J. Knigge, M. Sæther, & W. Waagen (Red.).

- Higher Education as Context for Music Pedagogy Research* (s. 19–40). Cappelen Damm Akademisk.
- Sandberg, R. (1996). *Musikundervisningens yttre villkor och inre liv. Några variationer över ett läroplansteoretiskt tema* [Doktorgradsavhandling]. Institutionen för pedagogik. Lärarhögskolan i Stockholm.
- Saunders, T. C. & Holahan, J. M. (1997). Criteria-specific rating scales in the evaluation of high school instrumental performance. *Journal of Research in Music Education*, 45(2), 259–272.
- Soep, E. (2004). Assessment and visual arts education. I E. W. Eisner & M. D. Day (Red.), *Handbook of research and policy in art education* (s. 579–583). Routledge.
- Stanley, M., Brooker, R. & Gilbert, R. (2002). Examiner perceptions of using criteria in music performance assessment. *Research Studies in Music Education*, 18(1), 43–52.
- Stobart, G. & Gipps, C. (1997). *Assessment: A teachers' guide to the issues*. Hodder & Stoughton.
- Swanwick, K. (1998). The perils and possibilities of assessment. *Research Studies in Music Education*, 10(1), 1–11.
- Thompson, S. & Williamon, A. (2003). Evaluating evaluation. Musical performance assessment as a research tool. *Music Perception*, 21(1), 21–41.
- Universitets- og høyskolerådet. (2015). *Retningslinjer for karakterskalaen bestått/ikke bestått* (Rapport). [https://www.uhr.no/\\_f/p1/iec171c06-38f2-4436-bce46621a6d32657/rapport\\_om\\_bestatt\\_ikkje\\_bestatt\\_10\\_112015.pdf](https://www.uhr.no/_f/p1/iec171c06-38f2-4436-bce46621a6d32657/rapport_om_bestatt_ikkje_bestatt_10_112015.pdf)
- UH-rådets utdanningsutvalg. (2016). *Generelle karakterbeskrivelser i UH-sektoren* (Rapport). <https://www.mn.uio.no/om/organisasjon/utvalg/studieutvalg/2018%20/karakterbeskrivelser-rapport-arbeidsgruppe-stut-28.11.18horing.pdf>
- UH-rådets utdanningsutvalg. (2009). *Tilleggsrapport fra arbeidsgruppe for å se nærmere på UH-sektorens generelle karakterbeskrivelser* (Rapport). <https://docplayer.me/36309836-Tilleggsrapport-fra-arbeidsgruppe-for-a-se-naermere-pa-uh-sektorens-generelle-karakterbeskrivelser.html>
- Vinge, J. (2011). Analytiske og holistiske tilganger til vurdering – en teoretisk drøfting av vurderingsstrategier og vurderingshandlinger i musikkfaget. I S.-E. Holgersen, S. G. Nielsen & L. Våkevä (Red.), *Nordisk musikkpedagogisk forskning årbok 13* (s. 199–220). Norges musikkhøgskole.
- Wibeck, V. (2010) *Fokusgrupper: om fokuserade gruppintervjuer som Undersøkningsmetod*. Lund, Studentlitteratur.

## VEDLEGG

### Karakterbeskrivelse for eksamen i utøvende musikk, bachelor (klassisk og kirkemusikk)

Karakter	Betegnelse	Beskrivelse av karaktertrinnene
A	Fremragende	Karakteren reserveres for kandidater som viser kunstnerisk kapasitet som klart utmerker seg. En fremragende prestasjon som på en overbevisende måte viser sammenheng mellom valg av musikalsk materiale, dramaturgi, stilforståelse og tolkning. Kandidaten demonstrerer et svært høyt teknisk nivå og overbevisende kommunikasjonsevner med medspillere og publikum.
B	Meget god	En prestasjon som viser stor grad av kunstnerisk kapasitet. En meget god framføring med en overbevisende kombinasjon av musikalsk bevissthet, instrumentbeherskelse og kommunikasjonsevne.
C	God	En prestasjon som viser god kunstnerisk kapasitet. En framføring med god musikalsk bevissthet, instrumentbeherskelse og kommunikasjonsevne. Det er ingen åpenbare mangler ved framføringen.
D	Nokså god	En prestasjon med noen mangler. Framføringen viser likevel samlet sett musikalsk bevissthet, tekniske ferdigheter og kommunikasjonsevne på et akseptabelt nivå.
E	Tilstrekkelig	En prestasjon der musikalsk bevissthet, instrumentbeherskelse og kommunikasjonsevne til sammen ikke tilfredsstillende mer enn det minimum som skal til for å vise at læringsmålene i emnet er oppnådd.
F	Ikke bestått	En prestasjon der musikalsk bevissthet, instrumentbeherskelse og kommunikasjonsevne er så svakt til stede at man ikke kan anse læringsmålene for oppnådd.

### Karakterbeskrivelse for eksamen i utøvende musikk, bachelor (jazz)

Bestått	Kandidaten kan uttrykke seg kunstnerisk, viser akseptabel form- og stilforståelse, formidlingsevne, gode improvisasjonsferdigheter og tilstrekkelige instrumentale og musikalske ferdigheter. En helhetlig konsertopplevelse der prestasjonen viser akseptabel oppfyllelse av vurderingskriteriene, og der man kan anse læringsmålene for nådd.
Ikke bestått	Kandidatens evne til å uttrykke seg kunstnerisk framstår som begrenset. Konsertopplevelsen som helhet, musikalsk bevissthet, formidling, improvisasjonsferdighet, musikalsk og instrumental mestring er ikke på et tilstrekkelig nivå. En prestasjon der oppfyllelsen av vurderingskriteriene er så svakt til stede at man ikke kan anse læringsutbyttene for oppnådd.