

Petter Erik Leirhaug

## **«Karakteren i seg selv gir jo ikke noe læring»**

En empirisk studie av vurdering for læring i kroppsøving ved  
seks videregående skoler i Norge

DOKTORGRADSAVHANDLING FRA NORGES IDRETTSHØGSKOLE • 2016

ISBN 978-82-502-0532-1

Petter Erik Leirhaug, NIH

## Sammendrag

Tema i denne artikkelbaserte avhandlingen er vurdering for læring (VfL) i kroppsøving. Både avhandlingens litteraturgjennomgang og tidligere internasjonale litteraturgjennomganger viser at selv om det har vært et økende tilfang av forskning på vurdering i kroppsøving de siste år, er det fortsatt få studier som har fokus på formative vurderingspraksiser og VfL.

I Norge ble læreplanreformen Kunnskapsløftet fra 2006 fulgt opp med en vurderingsforskrift og skoleutviklingsprosjekter som forutsatte implementering av formative vurderingspraksiser og VfL. Denne utdanningskontekst brukes i avhandlingen til å studere hvordan VfL forstås og implementeres i kroppsøving.

Til studien ble det rekruttert seks videregående skoler. Det empiriske grunnlaget er en undersøkelse blant alle elever og deres kroppsøvingslærere ved skolene. Ved hver av de seks skolene ble det gjennomført fokusgruppeintervju med de lærere som underviste i kroppsøving, mens datainnsamlingen blant elevene foregikk gjennom et digitalt spørreskjema om kroppsøving og vurdering. Det endelige utvalget teller 1486 elever og 23 kroppsøvingslærere.

Kombinasjonen av kvantitative og kvalitative metoder for innsamling og prosessering av empiri gjør at studien som helhet kan beskrives som et flermetodisk eller *mixed method* design. Den overordnede problemstillingen er belyst gjennom resultater fra fire studier som er publisert i artikkelform i engelskspråklige tidsskrifter.

Første artikkel, «Assessing with new eyes? Assessment for learning in Norwegian physical education», tar utgangspunkt i den norske konteksten der Kunnskapsløftet fra 2006 fungerer som eksempel på en læreplanreform med tydelig intensjon om å implementere VfL. Elev- og lærerdata analyseres og sammenlignes for å danne et bilde av hvordan ulike VfL strategier er en del av vurderingspraksisen ved de seks skolene. Resultatene viser liten grunn til å fokusere på forskjeller mellom skolene, men illustrerer at utdanningsmyndighetenes promotering av VfL foreløpig har ført til begrenset implementering og bruk av VfL strategier i kroppsøving. Majoriteten av lærerne uttrykker lite kjennskap til VfL og over halvparten av elevene rapporterer lite eller ingen erfaring med en eller flere av VfL strategiene. Ut fra en sammenligning av nøkkelstrategier for VfL identifiserer studien det å engasjere elevene mer direkte i vurderingsprosessene som et viktig satsingsområde.

Artikkel to, «Exploring the relationship between student grades and assessment for learning in Norwegian physical education», viser at elever som rapporterer mer erfaring med VfL i kroppsøving ikke får utbytte i form av bedre karakterer i faget sammenlignet med elever som rapporterer færre VfL-erfaringer. Funnet er ikke som forventet ut fra VfL teori, og i artikkelen diskuteres flere mulige forklaringer til dette. I tillegg diskuteres betydningen av andre variabler fordi sammenhengen mellom karakterresultat og elevenes rapporterte VfL-erfaring statistisk sett er liten. Selv om hovedfunnet er statistisk lite indikerer det at kroppsøving har validitetsproblemer både ved formative og summative vurderinger.

Tredje artikkel, «‘The grade alone provides no learning’: Investigating assessment literacy among Norwegian physical education teachers», utforsker kroppsøvlingslærernes vurderingskompetanse og beskrivelser av vurderingspraksis i lys av et teoretisk rammeverk for «assessment literacy» i kroppsøving. Resultatene demonstrerer et generelt behov for å utvikle «assessment literacy» blant kroppsøvlingslærerne, spesielt med fokus på dialog med elevene om vurdering, og kritisk refleksjon knyttet til vurdering og potensielt uheldige konsekvenser for elever. Artikkelen representerer også en empirisk utprøving av rammeverket for «assessment literacy» i kroppsøving og konkluderer med at begrepet kan danne et godt utgangspunkt for utvikling av forståelse og vurderingskompetanse hos lærere og i arbeid med kroppsøvlingslærerstudenter.

I avhandlingens siste artikkel, «‘It’s the other assessment that is the key’: Three Norwegian physical education teachers’ engagement (or not) with assessment for learning», presenteres tre lærere som selv og uoppfordret brukte VfL om noe de gjorde i sitt arbeid med vurdering i kroppsøving. De representerer lærere som oppfatter VfL som en mulig vei til å forbedre læringsarbeidet i kroppsøving. Ved å rekonstruere og diskutere lærernes vurderingshistorier illustreres hvordan de har forstått og praktiserer VfL i kroppsøving. VfL framstår som noe relativt nytt i lærernes praksis og historiene avdekker hvordan forståelsen og utforskningen av VfL infiltreres av tradisjonell vurderingskultur med fokus på summativ vurdering og karaktersetning. Dette kan se ut til å forsterkes gjennom vurderingsregelverkets krav til dokumentasjon og ved at læringsdiskursen i faget kroppsøving påvirkes av sterke idretts- og helsediskurser ellers i samfunnet.

Gjennom disse studiene og tilhørende kappe bidrar avhandlingen til å øke tilfanget av forskningsbasert kunnskap om faget kroppsøving i Norge generelt, og relatert til vurdering og VfL i kroppsøving spesielt. Artikkene inngår i et internasjonalt stadig voksende forskningsfokus på formativ vurdering og VfL i kroppsøving, men som resultatene illustrerer, er det, i alle fall i Norge, et stykke igjen før VfL er implementert i kroppsøving og kroppsøvlingsdidaktisk realisert på måter som gjør det relevant å undersøke betydning for faget og elevers læringsutbytte.

**Nøkkelord:** Vurdering for læring, formativ vurdering, vurderingspraksis, kroppsøving, implementering

## Abstract

The theme of this article-based thesis is assessment for learning (AfL) in physical education. The reviews of literature in this thesis and in previous international reviews show that although in recent years, there has been a growth of research on assessment in physical education, few studies focus on formative assessment practices and AfL.

In Norway, the 2006 school reform, known in English as the Knowledge Promotion Act, was followed by assessment regulations and projects for school development that required implementation of formative assessment practices and AfL. The intention of the thesis is to use the Norwegian context to examine how AfL is understood and implemented in physical education.

Six upper secondary schools were recruited to participate in the study. To collect data at each of the six schools, a digital questionnaire about physical education and assessment was administered to the students, while focus groups were conducted with the physical education teachers. The total sample counts 1,486 students and 23 physical education teachers. Because of the combination of quantitative and qualitative methods used for collecting and processing data the study can be described as having a *mixed method design*.

In addition to an introductory section (written in Norwegian), the thesis consists of four individual articles. Article one, "Assessing with new eyes? Assessment for learning in Norwegian physical education," sees the Norwegian 2006 school reform as an example of an education reform with a highly emphasised AfL agenda. Student data and teacher data were analysed and compared in order to investigate how assessment practices at the six schools reflect AfL key strategies. The results demonstrate that the educational authorities' emphasis on AfL has so far led only to limited implementation and use of AfL key strategies in physical education. The majority of the teachers expressed limited knowledge of AfL and over half of the students reported little or no experience with one or more AfL key strategies. Considering the comparison of the key strategies of AfL, this study recommends engaging students more directly in assessment processes as an important development area.

Article two, "Exploring the relationship between student grades and assessment for learning in Norwegian physical education," reveals that students who reported more experiences of AfL in physical education did not receive better grades in the subject than students who reported fewer AfL experiences. The result was not as expected according to AfL theory. While the main finding is statistically small, the study also discusses the influence of other variables, possible explanations for the results, and related implications. The main finding is statistically small, but indicates that physical education has problems with validity in both formative and summative assessments.

Article three, "'The grade alone provides no learning': Investigating assessment literacy among Norwegian physical education teachers," reports how physical education teachers reflected a proposed theoretical framework for assessment literacy in descriptions and discussions of their assessment practice in physical education. The results demonstrate a

general need to enhance assessment literacy among teachers, with a particular focus on critical engagement with assessment as well as involvement and dialogue with students. The article also represents an empirical test of the theoretical framework of assessment literacy for physical education teachers. It concludes that the framework can form a good basis for developing assessment understanding and assessment skills in physical education. The study highlights assessment as one of the areas that need to be developed in pre-service, as well as in-service, education of physical education teachers, and suggests that assessment literacy provides the framework and scope in which such work can take place.

The fourth article, “‘It’s the other assessment that is the key’: Three Norwegian physical education teachers’ engagement (or not) with assessment for learning,” reconstructs and shares assessment stories of three physical education teachers who used AfL unprompted to explain what they were practicing in respect to assessment in physical education. In this way, they are representative of teachers who believe in implementing AfL as a way to improve the learning activities in physical education. However, AfL emerges as something relatively new in teachers’ practices, and the assessment stories reveal how the teachers’ understanding and exploration of AfL are influenced by both a traditional focus on grading and new requirements for assessment documentation. The need for embedding AfL in learning theory may well be one of the strongest challenges to the meaningful implementation of AfL in physical education. Not only are most existing theories of learning defined cognitively, but the learning discourses in physical education are also, to a large extent, influenced by the powerful discourses of sport and related areas such as health.

The thesis, overall addresses the general lack of research-based knowledge on assessment in Norwegian physical education. Internationally, the articles represent a contribution to a growing research interest in formative assessments and AfL in physical education. The results, however, indicate that there is, at least in Norway, still a way to go before AfL are implemented as an integrated part of pedagogical practice in physical education, in ways that make it relevant to examine importance to the subject and student learning outcomes.

**Keywords:** Assessment for learning, formative assessment, assessment practices, physical education, implementation

**Translated title of the thesis:**

‘The grade alone provides no learning’: An empirical study of assessment for learning in physical education in six upper secondary schools in Norway

## Forord og takksigelser

I barneboka *Englebarn i tau* av Paul Leer-Salvesen fra 1994 er en stor del av handlingen knyttet til kroppsøving som fag. Forfatterstemmen kikker tilbake til da han var 12 år gammel og gikk i sjette klasse (i den niårige grunnskolen). Gutten blir kalt Preste-Paul på grunn av faren som er prest. Han klarer seg ganske bra på skolen, men «hater gym». Han er redd for kroppsøvingslæreren sin, «gymlærer Vang», og kommer ganske tidlig i boka med følgende advarsel: «Jeg tror alle husker gymlærerne sine hele resten av livet. Både de snille og de strenge. Vi drømmer om dem og forteller om dem og blir aldri helt ferdige med dem. Vi kan glemme hvem som var statsminister da vi gikk i sjette klasse, men ikke hvem vi hadde i gym.»

Når jeg har fått spørsmål om hva jeg skriver doktorgrad om, og jeg har svart «vurdering i kroppsøving», så har ikke det ført til løftede øyebryn, himlende øyne eller spørsmål om det er noe å skrive om. Tvert imot har reaksjonene tydet på at det oppfattes som et viktig tema. Neste steg er ofte at den som spurte begynner å berette fra egne, eller sine barns, erfaringer med kroppsøving, tidvis ledsaget av en halvartikulert forventning om at jeg skal si hva jeg mener om det berettede. Preste-Paul ser ut til å ha sine ord i behold. Kroppsøving kan føre til opplevelser som blir husket lenge. Faget ikke bare kan, men er ment å sette seg i kroppen og er noe alle virker å ha meninger om. I tillegg medførte innføringen av Kunnskapsløftet i 2006 endringer i grunnlaget for vurdering i kroppsøving som bidro til uro og usikkerhet blant mange kroppsøvingslærere og etter hvert tilspisset debatt om vurderingspraksisen i faget. I media kunne vi lese om urettferdighet, utstrakt bruk av testing, opplevd prestasjonspress og karakterjag i kroppsøving. Dels med dette som bakgrunn bestilte Kunnskapsdepartementet i 2011 en omfattende gjennomgang av faget og vurderingsordningen, og en ny, revidert læreplan ble tatt i bruk for kroppsøving fra høsten 2012.

Det er i denne tid og kontekst avhandlingen er blitt til. Jeg skal ikke legge skjul på at det har vært et privilegium å få arbeide med vurdering i kroppsøving i en periode da dette har vært gjenstand for debatt i ulike fagmiljøer så vel som på skolene og i media. Håpet er at mine møter med kroppsøvingslærere og andre underveis i arbeidet med avhandlingen, og eventuell videre bruk av resultatene, kan bidra til «å løfte faget» inn i framtiden. Uansett har jeg ikke gjort alt alene, og det er mange som fortjener en takk etter at jeg har levd med prosjektet i omtrent fem år.

Aller først vil jeg få takke *Seksjon for kroppsøving og pedagogikk* ved *Norges idrettshøgskole* for å tildele meg stipendiatet som gjorde denne avhandlingen mulig. Professor og seksjonsleder Gunn Engelsrud har hele veien vært positiv og oppmuntrende. Samarbeid med personer på seksjonen har vært min viktigste arena for å prøve ut og bryne fortolkninger og argumenter, og det har vært inspirerende at flere medstipendiater underveis har levert avhandlinger hvor kroppsøvingsfaget er sentralt. Ganske enkelt: Takk.

Kroppsøvingslærerne ved de seks skolene hvor datainnsamlingen foregikk fortjener en spesiell takk for imøtekommenhet og stor velvilje til å dele det jeg har opplevd som ekte omsorg for faget kroppsøving og dets potensielle framtider. I tillegg takker jeg ledelsen ved skolene for å tillate undersøkelsen og selvsagt alle elevene som fylte ut et relativt omfattende spørreskjema.

En takk rettes også til alle de lærere som på kurs og andre møter underveis har delt tanker og erfaringer, og med sine spørsmål rundt vurdering i kroppsøving har satt mine synspunkt på prøve. Videre takker jeg *Utdanningsdirektoratet* for å ha gitt meg muligheten til å bidra i arbeid med revisjon av læreplan i kroppsøving og utarbeidelse av veiledning til læreplanen. Spesielt vil jeg nevne de andre i faggruppen: Idar Lyngstad, Ingrid Nelvik og Lene Flagestad. Samarbeidet var moro og faglig givende.

Bibliotekene ved *Norges idrettshøgskole* og *Høgskulen i Sogn og Fjordane* har begge vært til uvurderlig hjelp. Jeg har også hatt gleden av å følge opplæringsprogrammet til NAFOL (Nasjonal forskerskole for lærerutdanning) hvor andre stipendiater, forelesere og kursholdere har bidratt med inspirerende diskusjoner og opptredt som eksemplariske kritiske venner når jeg har spurt om det. Ingen navn nevnt, ingen glemte.

Min veileder i avhandlingsarbeidet har vært professor Claes Annerstedt. Han har bidratt med konstruktive tilbakemeldinger og faglig dialog i alle prosjektets faser og vist evne til ekte veiledning ved å hjelpe fram viktige beslutninger når jeg har floket meg inn i det overivrige. Takk går også til personer tilknyttet *Department of Physical Education and Sport Sciences* ved *University of Limerick* i Irland for å ønske meg velkommen inn i faglige diskusjoner så vel som forelesninger og praktisk undervisning i kroppsøvingslærerutdanningen. Spesielt retter jeg en stor takk til min hovedkontakt og Head of Department, professor Ann MacPhail, som også er medforfatter på to av avhandlingens artikler, for inkluderende og faglig utfordrende samarbeid.

Sist, men ikke minst, vil jeg rette en varm takk til mine to nærmeste. Det er sagt at det er på barna en merker tiden. Ved prosjektets oppstart hadde sønnen i huset, Aasmund, ennå ikke begynt på ungdomsskolen. Ved avlevering er han elev i videregående skole. Både han og Mari har tålmodig, og etter hvert på ganske mange merkelige steder, måttet dele min oppmerksomhet med bøker, forskningsartikler og bærbar PC. Takk, Aasmund, du har vært en trofast påminner om at livet handler om meget annet enn formativ vurdering, regresjonsanalyser og hva kroppsøvingslærere sier og gjør. Takk, Mari, du har vært mitt viktigste faste punkt og støtte underveis, også ved de anledninger hvor du har gjort det klart at du var lei hele doktorgraden.

Kjøsnes i Fjaler & Oslo, 2016.

Petter Erik Leirhaug



## **Innhold**

(Sammendrag/Abstract)

<b>Forord og takksigelser</b> .....	7
<b>Innhold</b> .....	9
Tabeller .....	10
<b>Kapittel 1. Innledning</b> .....	11
Et styrket fokus på lærernes vurderingskompetanse .....	12
Å forstå og definere VfL .....	14
Forskning og kunnskap om VfL i kroppsøving .....	15
Erfaringer, avgrensninger og valg i norsk perspektiv .....	18
Formål og problemstillinger .....	20
Om innhold og form i avhandlingen .....	21
<b>Kapittel 2. Begreper, teori og kontekstualisering</b> .....	23
Vurdering og vurderingens formål .....	24
VfL i avhandling, teori og kroppsøving .....	25
Vurdering i kroppsøving, skolehistorisk skisse .....	30
Kroppsøving og vurdering i LK06 .....	33
Sammenlignbarhet og rettferdighet .....	35
Validitet og VfL i kroppsøving .....	38
Reliabilitet og gjennomsiktighet .....	41
«Assessment literacy» for kroppsøvingslærere .....	43
<b>Kapittel 3. Metodologiske og vitenskapsteoretiske detaljer, overveielser og valg</b> ....	45
Forskningsstrategi og design .....	45
En pragmatisk og sosiokulturell informert virkelighetsforståelse .....	46
Utvalg: Seks skoler, 1486 elever og 23 kroppsøvingslærere .....	52
Beskrivelse av lærerutvalget .....	54
Beskrivelse av elevutvalget .....	54

Fokusgruppene .....	56
Spørreundersøkelsen .....	58
Bruk av enveis ANOVA og regresjonsanalyse .....	60
Etiske overveielser .....	62
<b>Kapittel 4. Presentasjon av artiklene .....</b>	<b>64</b>
Artikkel 1: Assessing with new eyes? Assessment for learning in Norwegian physical education .....	64
Artikkel 2: Exploring the relationship between student grades and assessment for learning in Norwegian physical education .....	67
Artikkel 3: ‘The grade alone provides no learning’: Investigating assessment literacy among Norwegian physical education teachers .....	69
Artikkel 4: ‘It’s the other assessment that is the key’: Three Norwegian physical education teachers’ engagement (or not) with assessment for learning .....	71
<b>Kapittel 5. Sammenfatning og veier videre .....</b>	<b>74</b>
Avhandlingens bidrag og artiklenes sammenheng .....	74
Forskjeller mellom og i skolene .....	76
Noen hovedtrekk, utfordringer og begrensninger .....	77
Feedback i kroppsøving .....	82
Elevinvolvering og kritisk refleksjon .....	85
VfL i kroppsøving i Norge nå og i framtiden .....	88
<b>Kapittel 6. En empirisk epilog .....</b>	<b>92</b>
<b>Referanser .....</b>	<b>97</b>
<b>Artikler og vedlegg .....</b>	<b>113</b>
<b>Tabeller i kappen:</b>	
Tabell 1. Oversikt over utvalg av lærere og elever ved studiens seks skoler	
Tabell 2. Fordeling elevsvar mellom Vg1, Vg2 og Vg3	
Tabell 3. Fordeling svar på spørsmålet «hvilken oppvekstbakgrunn har du?»	
Tabell 4. Fordeling av rapporterte karakterer i kroppsøving ved siste halvårsvurdering	

## Kapittel 1. Innledning

«...i forhold til vurdering som fremmer læring, eller vurdering for læring, så prøvde jeg i løpet av fjoråret å prøve ut noen ting jeg ikke har gjort før, ved for eksempel at elevene har egenvurdering.»

Temaet for denne avhandlingen er vurdering for læring (VfL) i kroppsøving<sup>1</sup>. Utsagnet ovenfor er uttalt av en lærer som underviste i kroppsøving ved en av de seks videregående skolene som ble rekruttert til avhandlingens studie. Jeg har kalt læreren Ronald. Den tålmodige leser vil møte Ronald igjen senere, i kapittel 2, 4 og 6, og i en av avhandlingens artikler. Jeg har valgt å la et lærerutsagn innlede for å understreke avhandlingens empiriske fundament og fordi dette korte sitatet er egnet til å illustrere et gjennomgående trekk ved resultatene i avhandlingen.

Ronald er 59 år. Han har de siste tjue årene arbeidet som kroppsøvlingslærer på heltid, og deltok i et fokusgruppeintervju sammen med de fem andre lærerkollegene som underviste i kroppsøving ved samme skole. Ronald refererer til VfL for å beskrive noe han gjør i kroppsøving, og han forteller samtidig at deler av dette representerer noe nytt i hans arbeid med elevene. I sitatet viser han til egenvurdering, hvilket er et av de sentrale kjennetegn, prinsipper eller strategier som knyttes til formative vurderingspraksiser og VfL<sup>2</sup>. Ronald fremstår som en kroppsøvlingslærer med mange års erfaring som ganske nylig synes å ha oppdaget noe han finner verd å utforske gjennom utprøving i egen praksis.

Hvor utbredt en slik utforskning eller bruk av VfL er i kroppsøving, og hva det kan innebære for læring og elever, er noe av hva denne avhandlingen søker å gi svar på. Men i avhandlingens oppspill er poenget at Ronald, ved å referere til VfL for å beskrive deler av egen praksis, plasserer seg i sentrum av de siste par tiårs dramatiske endring av diskursen

---

<sup>1</sup> Kroppsøving brukes i denne avhandlingen om skolefaget kroppsøving. Kroppsøving kan også forstås mer generelt som «øvelser for kroppen», fysisk fostring og beslektede begreper som peker i retning av alle bidrag til kroppslig læring og utvikling. Denne doble forståelsen deler kroppsøving med de engelskspråklige land hvor faget kalles «physical education». I Sverige heter tilsvarende skolefag «idrott och hälsa», i Danmark «idræt» og rundt om i verden finner vi ulike engelskspråklige varianter som «health and physical education» og «human movement education». I avhandlingen bruker jeg det norske «kroppsøving» inkluderende som fellesbenevnelse for de ulike variantene av skolefaget. Dette tilsvarer bruken av «physical education» som skolefagets fellesbetegnelse i de engelske artiklene. Der kroppsøving avgrenses spesifikt til å gjelde faget i Norge vil dette framgå av sammenhengen.

<sup>2</sup> Teoretisk rammeverk for vurdering for læring (VfL) behandles i avhandlingens kapittel 2. Der finnes en oversikt over nøkkelstrategier basert på teoriutviklingen gjort av Black og Wiliam (1998a, 1998b, 2006, 2009, 2012a). Det å involvere elevene i vurdering av eget arbeid er med i Utdanningsdirektoratets (2008) kjennetegn på vurdering for læring, og det er nedfelt et formelt krav om en slik praksis i Forskrift til opplæringslova (2006) § 3-12.

omkring vurderingens rolle i læring og utdanning. Spesielt har fokus og forskning på formativ vurdering, eller vurderingens læringsfremmende funksjoner, informert og vært med å danne denne reevaluering av vurderingens pedagogiske betydning (Black & Wiliam, 1998a, 1998b, 2006, 2009, 2012a; Gardner, 2012; Hattie & Timperley, 2007; Klenowski & Wyatt-Smith, 2014; Sadler, 1989, 1998; Stobart, 2008; Wiliam, 2010, 2011). VfL framtrer utover 1990-åra som et konsept for formativ vurderingspraksis, og Wiliam (2011) bemerker i sin oppsummering av utviklingen at VfL virker å bli foretrukket framfor begrepet formativ vurdering. En grunn til dette kan være, som Pryor og Crossouard (2008) skriver, at VfL synes å kommunisere bedre inn i skolekontekstene. Beskrivelsene er treffende for hvordan VfL er det begrep som virker å ha funnet gjenklang og fått fotfeste i norsk utdanningskontekst (Dobson & Engh, 2010; Engh, Dobson & Høihilder, 2007; Slemmen, 2010; Utdanningsdirektoratet, 2008, 2011).

### **Et styrket fokus på lærernes vurderingskompetanse**

Vurderingsregelverket som fulgte skolereformen Kunnskapsløftet i 2006 (LK06) bygger på og henviser til forskning som viser at vektlegging av formativ vurdering og VfL bærer et betydelig potensiale til å støtte elevenes faglige læring og øke læringsutbyttet (Slemmen, 2010; Tveit, 2009, 2014; Utdanningsdirektoratet 2008, 2011; Wilson, 2014). Evalueringene av skolereformene på 1990-tallet viste at norske lærere var mer opptatt av aktivitetene i klasserommet enn hva elevene skal lære (Haug, 2003). Det ble pekt på mangelfull læreplananalyse og planlegging rettet mot elevenes tilegnelse av ferdigheter gjennom frie arbeidsmåter. Tilbakemeldinger i klasserommene preges av ros for innsats og prestasjoner, og knyttes sjelden til hvordan elevene skal forbedre sitt læringsresultat (Klette, 2003). På bakgrunn av dette tilhører vektleggingen av formativ vurdering en politisk visjon om å skape en kultur for læring (NOU 2002:10, 2002; St.meld. nr. 30 (2003-2004), 2004). I reformen får dette uttrykk som undervisvurdering med et klart artikulert mål om å endre et tradisjonelt fokus på vurdering av læring<sup>3</sup>, og summativ vurdering i form av karaktersetning, i retning av VfL (Utdanningsdirektoratet, 2008, 2011). Dette følges opp i stortingsmeldingen *På rett vei: Kvalitet og mangfold i fellesskolen* (Meld St. nr. 20 (2012-2013), 2013) der *Kunnskapsdepartementet* ønsker å videreføre nasjonal satsing på VfL for å heve lærernes vurderingskompetanse og styrke sammenhengen mellom undervis- og sluttvurdering.

---

<sup>3</sup> Distinksjonen mellom vurdering av læring og VfL tilskrives i mange publikasjoner Gipps (1994), men – som det vises til i kappens kapittel 2 – Harry Black (1986) var tidligere ute. I enkelte amerikanske artikler nevnes Stiggins (2002) som opphav til VfL. Avhandlingens begrepsbruk behandles nærmere i kapittel 2.

Samtidig var reformen LK06 dels motivert av at resultatene i undersøkelser som PISA, TIMSS og PIRLS<sup>4</sup> ikke svarte til forventningene når det gjelder norske skoleelevers læringsresultat (Dale, 2010; Telhaug, 2005), og er understøttet av en offentlig forvaltningstenkning preget av nyliberalisme og som favoriserer det vi kan kalle accountability-perspektivet (Aasen, 2012; Møller & Skedsmo, 2013). Tveit (2014) konstaterer at den norske politiske vurderingsdiskursen herjes av uavklarte spenninger mellom summative og formative formål med vurdering. Dette klimaet for læreplan- og vurderingsreformer er en delt internasjonal erfaring, og representerer en utfordring for implementering av formative vurderingspraksiser (Berry & Adamson, 2011; Birenbaum m.fl., 2015). Utfordringer med utdanningsmyndigheters promotering av VfL innenfor systemer som samtidig setter krav til dokumentasjon og summative sluttvurderinger er også synliggjort i kroppsvøvingkontekst (Green, 2008; Seger, 2014; Tolgfors, 2014).

Endringene i synet på vurdering gjenspeiler de sosial-konstruktivistiske kunnskapsparadigmer og pedagogiske teorier som vokser fram mot slutten av det tjuende århundre, der eleven og læringen, eller den lærende, settes i sentrum (Dale & Wærness, 2006; Dobsen, Engh & Smith, 2009; Lundahl, 2011; Shepard, 2000). I kontrast til tradisjonelle forestillinger om vurdering understøttet av paradigmer for vitenskapelig måling, sosial effektivitet og behavioristiske læringsteorier, hvor det nærmest var et ideal at vurdering av elevens faglige kompetanse kunne gjøres uavhengig og atskilt fra opplæringen i faget, understreker de nyere perspektiver på vurdering hvordan vurdering, undervisning, læreplan og læringsteori henger tett sammen (Hay & Penney, 2013; James, 2006, 2012; Shepard, 2000; Stobart, 2008).

Når jeg ovenfor kalte denne endringen dramatisk, var det ikke bare i lys av utvikling i pedagogikk og fagdidaktikk, men også ut fra personlig erfaring. I dag arbeider jeg i lærerutdanning og vurdering inngår der som et selvsagt og sentralt tema. Derimot kan jeg ikke huske at vurdering var et tema i pedagogikkundervisningen da jeg studerte til å bli lærer rundt 1990. Eneste relevante minne fra studietiden er møtet med karaktersetning som utfordring under skolepraksis i ungdomsskolen. Vi, en gruppe lærerstudenter i praksis, skulle foreslå karakter og kommentar på en tradisjonell prøve vi gjennomførte med elevene, for deretter å diskutere våre vurderinger med elevenes faglærer, som i dette tilfellet også var vår øvingslærer. I etterrefleksjonens lys, kvalifiserer dette som et klassisk eksempel på

---

<sup>4</sup> Forkortningene viser til tre overnasjonale sammenligningsstudier hvor Norge er med: PISA (Programme for International Student Assessment) studerer ulike sider ved utdanningskvaliteten i OECD-landene, TIMSS (Trends in International Mathematics and Science Study) måler kompetanse i matematikk og naturfag, mens PIRLS (Progress in International Reading Literacy Study) gjelder leseferdighetene til elevene.

tradisjonell vurdering, designet med det formål å måle elevenes prestasjoner på slutten av et pedagogisk forløp, hva og hvor mye eleven kan vise av det han/hun skulle ha lært. I henhold til formålet er eksemplet betegnende for vurdering av læring, og kontrasterer VfL som tema for denne avhandlingen.

Pophams (2011) bekjennelse av ikke å gi nok oppmerksomhet til betydningen av vurderingskompetanse i hans tidlige år som lærerutdanner, kan sies å tilføre min erfaring fra egen lærerutdanningen en internasjonal dimensjon. Fra at Popham knapt erindrer å ha undervist i vurdering i sin tidlige fase som lærerutdanner, inngår det nå i hans praksis som en nødvendig og integrert del i arbeidet med å kvalifisere nye lærere. Mens endringen er uttrykt og tydelig i tilfellet Popham (2011), tyder rapporter fra lærerutdanningen i Norge på at det fortsatt er et stykke igjen før norske lærerutdannere besitter tilstrekkelig kompetanse til å utvikle VfL kompetente lærere for fremtiden (Arnesen m.fl., 2010; Engelsen & Smith, 2010). Artikkel 3 i avhandlingen (Leirhaug, MacPhail & Annerstedt, 2016) antyder at utfordringen kan være større i tilknytning til kroppsøving sammenlignet med andre fag, og Backman og Larsson (2013) viser konkret til vurdering som et av de områdene som må forbedres og samtidstilpasses i nordiske kroppsøvingslærerutdanninger<sup>5</sup>.

### **Å forstå og definere VfL**

Det ligger i selve formuleringen VfL at det i vid forstand menes all vurdering som har til hensikt å støtte og øke elevenes læring. Dette skiller VfL fra begrepet formativ vurdering fordi også summative vurderinger kan benyttes og iverksettes med læring som primært formål. Brookhart (2007) viser hvordan formativ vurdering har utviklet seg til å omfatte en rekke ulike praksiser og følgelig kan forstås på forskjellige måter. «Assessment reform group» (ARG) var en gruppe bestående av vurderingsforskere<sup>6</sup> i England som rundt århundreskiftet argumenterte for å bruke VfL heller enn formativ vurdering og denne avhandlingen legger deres ofte siterte definisjon til grunn for forståelsen av VfL:

---

<sup>5</sup> Litt kuriøst i denne sammenheng kan det nevnes at Tindall og Enright (2013) ikke har vurdering blant sine overveielser når de tar opp spørsmålet «what do physical education teachers need to know?» som en del av samtidens kroppsøvingsdebatt.

<sup>6</sup> ARG ble oppløst i 2010. Inntil da hadde følgende forskere i perioder vært medlemmer: Jo-Anne Baird, Paul Black, Patricia Broadfoot, Richard Daugherty, Bryan Dockrell, Kathryn Ecclestone, John Gardner, Caroline Gipps, Wynne Harlen, Louise Hayward, Mary James, Paul Newton, Desmond Nuttall, Judy Sebba, Gordon Stobart og Dylan Wiliam.

the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there (Assessment Reform Group, 2002, s. 2)<sup>7</sup>.

Forstått på denne måten handler VfL om vurdering som foregår der læring skjer og utgjør en del av det daglige og kontinuerlige arbeidet blant lærere og elever. Internasjonalt debatterer forskningslitteraturen fortsatt betydningen av VfL og under hvilke forutsetninger de potensielle læringseffekter av denne type vurderingspraksis kan realiseres for elevene (Baird m.fl., 2015; Bennett, 2011; Dunn & Mulvenon, 2009; Wiliam, 2011). Det virker likevel å være bred enighet om at VfL og formativ vurdering kan beskrives og gjenkjennes gjennom noen sentrale prinsipper<sup>8</sup> som handler om at elevene forstår læringsmålene, er kjent med kriterier for vurdering og får tilbakemeldinger som fører i retning av læringsmålene. Videre tilhører det prinsippene at elever engasjeres i å vurdere eget arbeid og at undervisning justeres slik at den tar hensyn til mulig utbytte av VfL. I artiklenes studier operasjonaliseres VfL ut fra noen slike prinsipper, kjennetegn eller nøkkelstrategier. Dette utdypes i avhandlingskappens kapittel 2, og der vises også hvordan spørsmål rundt sammenlignbarhet, validitet og sosial rettferdighet vil være gyldige for VfL og andre alternativer til tradisjonell summativ vurdering, men at de på grunn av det læringsfremmende formålet må møtes på andre måter og med fornyede verktøy (Gardner, 2012; Klenowski & Wyatt Smith, 2014).

### **Forskning og kunnskap om VfL i kroppsøving**

Vurdering har historisk ikke fått mye oppmerksomhet i kroppsøvingforskning (Carrol, 1994; Georgakis & Wilson, 2012; Hay & Penney, 2013). I litteraturgjennomgangen Hay (2006) gjennomførte for *The Handbook of Physical Education* (Kirk, Macdonald & O'Sullivan, 2006) bemerket han at det, i lys av det økende fokus på vurdering i pedagogisk litteratur generelt, var «somewhat strange that assessment in physical education has been neglected by

---

<sup>7</sup> Med utgangspunkt i utviklingen av teorien rundt formativ vurdering reformulerte Black og Wiliam i 2009 en definisjon som de mente også ivaretok denne definisjonen fra Assessment Reform Group. Den utdyper perspektivet noe, er kompatibel med forståelsen av VfL i denne avhandlingen og gjengis derfor her: Vurdering fungerer formativt “to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.” (Black & Wiliam, 2009, s. 9).

<sup>8</sup> I tilknytning til VfL brukes begreper som prinsipper, strategier, indikatorer og kjennetegn om hverandre i litteraturen, både på norsk og engelsk (se også note 2). Det snakkes også om VfL nøkkelprinsipper og VfL nøkkelstrategier. I denne avhandlingen likestilles slike omtaler og vil benyttes åpent om hverandre. I engelskspråklig litteratur er det mest vanlig å møte formuleringene «key strategies» eller «key principles», hvorav begge benyttes i avhandlingens artikler.

academic researchers» (s. 313)<sup>9</sup>. Senere gjennomganger har rapportert en tydelig økning i forskning på vurdering i kroppsøving (Chan, Hay & Tinning, 2011; Hay & Penney, 2013; López-Pastor m.fl., 2013), men sier samtidig at «most of the studies do not specifically explore the pedagogic work of assessment or its impact on learning in PE» (Chan, Hay & Tinning, 2011, s. 3). I en eksplorerende undersøkelse med direkte fokus på VfL i kroppsøving og skoleidrett, pekte Georgakis og Wilson (2012) på en tendens i både forskningslitteratur og mer instruksjonsorienterte bøker «to simply reiterate to the reader that assessment need to be ‘authentic’» (s. 41), og konstaterte «a disappointing lack of practical examples» (s. 42). Deretter intervjuet de 17 lærere i kroppsøving og skoleidrett i Australia for å lære mer om vurderingspraksis og tenkning, og deres konklusjon taler for seg selv:

Although we started this study with the intention to explore *assessment for learning* it was clear from the outset that PESS [Physical education and school sport] teachers did not use this term or often discuss links between assessment and learning. Our inductive coding of interview transcripts did not produce any categories of discussion on AfL. (Georgakis & Wilson, 2012, s. 49).

Georgakis og Wilsons (2012) observasjon og erfaringer finner støtte i en relativt fersk review av forskningslitteratur om (alternativ) vurdering i kroppsøving gjennomført av López-Pastor m.fl. (2013). Litteraturgjennomgangen deres viser det samme som resultatene av mine systematiske søk via databaser<sup>10</sup>: Empirisk forskning på formativ vurdering i faget kroppsøving er ganske sjelden. Siden de første slike studier dukket opp rundt 1990, er det ikke flere publisert i engelskspråklige «peer-review journals» enn at det er overkommelig å lese alle om en har litt tid til overs.

López-Pastor m.fl. (2013) bruker begrepet «alternative assessment» som fellesbetegnelse for ulike alternativer til mer tradisjonelle vurderingsformer i kroppsøving.

---

<sup>9</sup> I en kunnskapsoversikt over kroppsøvningsforskning i Norge for perioden 1978 til 2010 lar Jonskås (2010) seg overraske tilsvarende. Hun finner at det er «lite forskning på høyere nivå, særlig når det kommer til elevvurdering og kroppsøving generelt. Det er svært mangelfull kunnskap om hva elever lærer og hvordan undervisningen foregår» (s. 7). Med høyere nivå refererer hun til de krav som stilles til Universitet- og høyskolesektoren ved vitenskapelig publisering.

<sup>10</sup> I august 2011 gjorde jeg testsøk for litteratur i noen databaser med søkeordene «PE» og «physical education» koplet til henholdsvis «assessment»/«grading» i tittel. Resultatet var ikke overveldende. *SPORTDiscus* kom ut med flest treff, henholdsvis 286/41, *SAGE* hadde 6/0, *ERIC* 78/11, *SpringerLink* 2/0 og *Academic Search Premier* 46/6. Tilsvarende søk i august 2015 gav med bibliotekenes søketjeneste *Oria* resultat 396/48 og 301/46 i *SPORTDiscus*. Ved nærmere ettersyn handler mange av artiklene om «physical assessment», og således ikke om pedagogisk vurdering. De fleste aktuelle artikler er inkludert i reviewen til López-Pastor m.fl. (2013) og jeg har derfor brukt denne som utgangspunkt i kunnskapskartlegningen.



De alternative vurderingsformene figurerer under ulike navn som formativ vurdering, autentisk vurdering, VfL og integrert vurdering. Reviewen omtaler 26 publikasjoner med denne type «alternative assessment» i kroppsøving for perioden 1988-2011, av disse er bare ti empiriske studier i betydningen at de har gjennomført undersøkelser i skolehverdagen eller for eksempel blant kroppsøvingslærere<sup>11</sup>. Etter år 2000 er det fem slike empiriske studier hvor ikke den alternative vurderingsformen er gjennomført som et ledd i forskningsprosjektet. Av disse vurderer jeg tre til å ha et annet hovedfokus enn selve vurderingen (Hay, 2010; Hay & Macdonald, 2010a, 2010b), hvilket levner to studier nyere enn 2000 som faktisk undersøker alternativ vurderingspraksis i ordinær kroppsøving (Chen, 2005; Mintah, 2003). Verd å notere er at både Chen (2005) og Mintah (2003) fant formative vurderingspraksiser i sine studier av kroppsøving, men at disse praksisene ikke framtrer i en tydelig pedagogisk eller teoretisk kontekstualisering.

Selv ved å inkludere de studier fra reviewen som er publisert på engelsk og bruker empiri fra skolefaget kroppsøving der VfL forekommer som del av en intervensjon, eller på andre måter er forskerinitiert, står det klart at kunnskapen om VfL i sammenheng med kroppsøving ikke er omfattende. Og selv om de få studiene som finnes viser hvordan kroppsøvingslærerne opplever utfordringer med å ta i bruk VfL, spesielt i forhold til tidsbruk, indikerer de positivt utkomme både for undervisning i kroppsøving og elevenes læringsferfaringer (Chan, Hay & Tinning 2011; MacPhail & Halbert, 2010; Ní Chróinín & Cosgrave, 2013). Dette er likevel å betrakte som underdokumentert og er dels et resultat av at de fleste empiriske studier av vurdering i kroppsøving altså ikke tar for seg vurderingens potensielle «pedagogical work» (Chan, Hay & Tinning, 2011; Hay, Tinning & Engestrom, 2015), men fokuserer heller validitetsutfordringer som at lærere ikke klarer å redegjøre for hvilke kriterier de faktisk bruker for å vurdere elevenes fagkompetanse (f. eks. Hay & Macdonald, 2008; Annerstedt & Larsson, 2010; Svennberg, Meckbach & Redelius, 2014) og manglende samsvar mellom offisiell læreplan og, det som av elever oppfattes som, eller av

---

<sup>11</sup> Mange av senere års publikasjoner internasjonalt på vurdering i kroppsøving er teoristudier i betydningen at de ikke presenterer egen empiri. Majoriteten er av typen som søker å besvare noen av de validitets- og rettferdighetsutfordringene vurderingsforskningen har pekt på. Thornburn (2007) og Melograno (2007) diskuterer hvordan sluttvurdering og/eller eksamen i kroppsøving har kommet i utakt med fagets utvikling og hvordan dette kan forbedres, noe Thornburn (2008) for eksempel forsøker gjennom å trekke inn fenomenologiske perspektiver i større grad. Disse studiene handler eksplisitt om å bedre kvaliteten i vurdering av læring. Harvey og van der Mars (2010) og van der Mars og Harvey (2010) gjør et fagdidaktisk arbeid med tilrettelegging av aktivitet for ferdighetslæring og vurdering av aktiviteten i skolekontekst, et progressivt arbeid som har perspektiv på vurdering av og for læring. Det samme gjelder Hay og Penneys (2009) forslag til hvordan kroppsøving på tvers av landegrensene og læreplaner med utgangspunkt i Bernsteins (1975) «three message systems for schooling» kan få et redskap til å analysere og/eller bedre kvaliteten i vurderingsarbeidet.

lærere er uttrykte, vurderingskriterier (f. eks. Redelius & Hay, 2009, 2012; Young, 2011; Zhu, 2015). I tillegg viser studier at det heller ikke nødvendigvis er slik at kroppsøvingslærere faktisk vurderer ut fra de kriterier de hevder, og kanskje tror, de bruker (Chatzopoulos, Erdmann & Tsormbatzoudis, 2006). Noen senere studier har med et sosiokulturelt perspektiv sett på hvordan vurdering kan påvirke og blir påvirket av elevene (f.eks. Hay, 2010; James, Griffin & Dodds, 2009; Redelius, Fagrell & Larsson, 2009), og disse studiene minner oss om at implementering av VfL i kroppsøving ikke er en løsning, men vil utgjøre en av flere stemmer i fagforståelsens konstruksjon. Implementering av VfL må ledsages av en kritisk dialog om hvilken kunnskap og kompetanse som verdsettes, om distribusjon av makt og sosial rettferdighet. Tolgfors og Öhman (2015) viser hvordan fokus på VfL nøkkelstrategier i kroppsøving like gjerne kan brukes som et redskap til å styre og kontrollere elever som til å fremme deres læring. I sin masteroppgave om vurdering sett med elevøyne i norske klassegrupper fant Palm (2014) praktisk talt et fravær av slike kritiske perspektiver på vurdering generelt og VfL spesielt.

### **Erfaringer, avgrensning og valg i norsk perspektiv**

Min inngang til å studere VfL i kroppsøving stammer likevel ikke først fra den ovenfor beskrevne forsknings- og kunnskapsmangel på området. Spørsmålene rundt VfL ble aktualisert mens jeg var kroppsøvingslærer og startet som en kombinasjon av impulser fra innføringen av LK06 og det jeg med Macdonald (2002) vil kalle «practice informed theory». Jeg forstår dette nært beslektet med Handal og Lauvås (1999) sitt begrep om praksisteori, og det handler om hvordan læreren allerede står i og har en praksis som vil prege fortolkning av ny teori og ha betydning for hvordan teorien eventuelt inkorporeres i lærerens praksis.

Altså, ved innføringen av kunnskapsløftet i 2006 var jeg lærer i en videregående skole hvor jeg underviste i kroppsøving og et par andre fag. Ledelsen ved skolen satte av noe tid til hvert fag for at faglærerne sammen skulle arbeide med nye læreplaner og utarbeide kjennetegn på måloppnåelse. Mine kroppsøvingslærerkolleger og jeg fant raskt ut at dette handlet om mer enn kjennetegn på måloppnåelse. Etter å ha satt oss inn i de nye vurderingsforskriftene, med innspill fra vurderingslitteratur<sup>12</sup>, begynte vi å jobbe med hvordan vi kunne planlegge bedre tilbakemeldinger til elevene, dokumentere læring på meningsfulle måter og, som Ronald, engasjere elevene i egenvurdering. Vi startet i det små

---

<sup>12</sup> Vi leste blant annet *Inside the black box* av Black og Wiliam (1998b) og disse leseerfaringene prøvde vi så å omplante i egen praksis, ikke ulikt de erfaringer med avgjørende fagtekster som er nevnt hos Casey (2013). Noen av mine refleksjoner om vurdering i kroppsøving etter første år med LK06 finnes i Leirhaug (2008).

og høstet erfaringer underveis. Allerede det første året hadde vi revidert vår tenkning flere ganger, og jeg er redd noen elever var ganske forvirret. Men selv i lys av denne premature utforskning av praksis med VfL i kroppsøving mente jeg å se positiv betydning for læringsutbytte, og tross noe forvirring virket de fleste elevene fornøyd med opplegget. Dette var en av de viktigste erfaringene jeg tok med meg da jeg begynte å jobbe i lærerutdanning to år senere. Mine egne erfaringer fra å undervise i kroppsøving i videregående skole har vært viktig også i utformingen av prosjektet for denne avhandlingen.

Forskningslitteratur viser at ikke alle delte min relativt positive erfaring. I sin masteroppgave intervjuet Bomo (2008) kroppsøvingslærere på Vg1 om synspunkt på og erfaringer med implementeringen av Kunnskapsløftet. Han identifiserte endringer i vurderingsordning som en sentral kilde til usikkerhet og frustrasjon blant lærerne. At lærerne reagerte veldig forskjellig på den nye reformen er godt dokumentert i de studier av vurdering i kroppsøving som er utført etter 2006 (Arnesen, Nilsen & Leirhaug, 2013; Engvik, 2013; Prøitz & Borgen, 2010; Sandvik m.fl., 2012; Vinje, 2008a)<sup>13</sup>. Sandvik m.fl. (2012) undersøkte formative vurderingsprosesser i fire fag og sier at i «et validitetsperspektiv fremstår altså kroppsøving [som] et minefelt av motstridende føringer og intensjoner» (s. 153). Selv om forskning på vurdering i kroppsøving totalt ikke har vært omfattende i Norge (Jonkås, 2010, 2011), har debatten om vurderings spørsmål vært opphetet, spesielt i kjølvannet av den nye læreplanen og vurderingsregelverket som ble innført i 2006. Usikkerhetene om vurderings spørsmål i kroppsøving var så alvorlig at *Utdanningsdirektoratet* på oppdrag fra *Kunnskapsdepartementet* gjennomførte en grundig gjennomgang av faget (Lyngstad m.fl., 2011). Dette resulterte i en revidert læreplan i kroppsøving med iverksettelse fra august 2012<sup>14</sup> (Utdanningsdirektoratet, 2012).

Det er dette som utgjør bakteppet når det i tredje og fjerde artikkel i avhandlingen (Leirhaug, MacPhail & Annerstedt, 2016; Leirhaug & MacPhail, 2015) påpekes at det har vært andre utfordringer ved vurdering enn det å utvikle god og effektiv VfL som har opptatt kroppsøvingslærerne i Norge etter LK06. Endringer rundt vurderingsgrunnlaget knyttet til

---

<sup>13</sup> Det er også gjort studier på masternivå, men disse er ikke inkludert her da de er av svært ulik kvalitet. I utgangspunktet deler jeg synet til Borgen og Engelsrud (2015) om at «masterstudentene må betraktes som kunnskapsprodusenter for feltet» (s. 66), og master- og hovedfagsoppgaver med relevans for avhandlingens tema er i tråd med dette inkludert der de er vurdert til å bidra til å belyse funn og diskusjoner.

<sup>14</sup> Datamaterialet i denne avhandlingen er innsamlet før revidert læreplan ble iverksatt og relateres derfor til læreplanen før revisjon av august 2012. Dette er ikke gjort et poeng utav i de engelske artikkelene, men vil likevel bli adressert i kappens kapittel 5. Fokus vil da være på hva resultatene i avhandlingen kan bety i lys av revidert læreplan. Antisipierte kommentar er – i likhet med at deler av VfL er gyldig på tvers av kontekster – at synet på betydningen av formativ vurdering ikke ble vesentlig endret med revisjonen. Det var en tydeliggjøring av fagets formål med oppfølgende justeringer av kompetanseformuleringer og særformuleringen om «elevens innsats som del av vurderingsgrunnlaget» i Forskrift til opplæringsloven (2006) § 3-3 som utgjorde hovedendringene.

innsats og hensyn til elevenes forutsetninger, samt diskusjoner rundt bruk av ulike tester som grunnlag for karaktersetning, rettferdighet og stort sprik i vurderingspraksis mellom skoler, har preget diskusjonen både i faglige forum og i media (Arnesen, Leirhaug & Nilsen, 2010; Elnan & Sando, 2014; Vinje, 2008b; Vinje & Brattenborg, 2016). En vesentlig utfordring for faglig utvikling i kroppsøving, og her spesielt vurdering i kroppsøving, er at fire av ti lærere som underviser i faget ikke har noen formell kroppsøvingsfaglig utdanning (Lagerstrøm, 2007; Lagerstrøm, Moafi & Revold, 2014; Leirhaug & Midthaugen, 2014). Ut fra dette valgte jeg å rekruttere videregående skoler til avhandlingsstudien. På videregående trinn i opplæringen er det oftest kroppsøvingslærere med fagutdanning, noe jeg vurderte som vesentlig med tanke på interesse for å utvikle faget og bidra med fagdidaktisk refleksjon relatert til VfL. Videre har videregående skole mer etablerte nettverk på tvers av skoler innenfor fylket sammenlignet med hva tilfellet er i grunnskolen. Derfor kunne jeg kanskje forvente at de lokale kulturene som var påpekt i følgeforskningen for Kunnskapsløftet (Møller, Prøitz & Aasen, 2009) var mer moderert i videregående skole enn på ungdomsskoler.

### **Formål og problemstillinger**

Formålet med avhandlingen er å studere implementering og forståelse av VfL i kroppsøving i perspektiv fra lærere og elever ved samme skoler, og gjennom dette å bidra til den generelle forståelse av og teoriutvikling omkring VfL i kroppsøving. Avhandlingen representerer således et ønske om å produsere ny kunnskap som kan være med å belyse de komplekse sammenhengene mellom undervisning, vurdering og læring i kroppsøving. Med utgangspunkt i avhandlingens formål, kunnskapsløftets vektlegging av undervisvurdering og valg av videregående skole som inngang er følgende problemstilling formulert:

*Hvordan forstås og implementeres vurdering for læring i kroppsøving i videregående skole?*

Implementering behandles som et åpent begrep for å studere oppfattelse og utbredelse av VfL sett fra elev- og lærerperspektiv, mens forståelse i første rekke handler om kroppsøvingslærernes kunnskap om VfL og beskrivelser av relevante praksiser. Den overordnede problemstillingen har blitt utforsket gjennom fire studier som alle er ledet av en mer spesifikk hypotese eller forskningsspørsmål. Disse gjengis her i engelsk språkdrakt, identisk med formuleringene slik de forefinnes i avhandlingens artikler:

Artikkel 1: *How do students' and teachers' perspectives of assessment practices in physical education reflect AfL<sup>15</sup> key principles?*

Artikkel 2: Hypotese: *There is a positive relationship between student-reported higher levels of AfL experiences in physical education and better grades in the subject.*

Artikkel 3: *How do Norwegian physical education teachers reflect assessment literacy in descriptions and discussions of their assessment practice?*

Artikkel 4: *How is AfL motivated, understood and enacted by physical education teachers who are trying to integrate AfL into their teaching and assessment practice?*

Undersøkelsen av de fire forskningsspørsmålene er basert på empiriske data innsamlet via (i) et internettbasert spørreskjema om kroppsøving og vurdering besvart av elever ved seks videregående skoler og (ii) seks fokusgruppeintervju som inkluderte de lærere som underviste i kroppsøving ved de samme skolene. Kombinasjonen av kvantitative og kvalitative metoder for innsamling og prosessering av data gjør at studien som helhet kan beskrives som et *mixed method design* (Creswell, 2009; Tashakkori & Teddlie, 2003; se også kapittel 3). Den overordnede problemstillingen vil belyses gjennom en sammenfattende diskusjon av resultatene i de fire artiklene (kapittel 5).

### **Om innhold og form i avhandlingen**

Avhandlingen består av fire artikler og en sammenbindingstekst. I tråd med det som er vanlig i norske artikkelbaserte avhandlinger omtales den sammenbindende teksten som kappen. Ut over å diskutere den overordnede problemstillingen, gir kappen i hovedsak en utdypning av prosjektets teoretiske og metodiske rammeverk, samt plasserer artiklene i en norsk utdanningskontekst og diskuterer resultatene i lys av dette.

Artiklene er skrevet på engelsk med mål om å bidra inn i den internasjonale diskurs og utvikling av forskningsfeltet vurdering i kroppsøving. Ut over at det fortsatt er sparsomt med forskning på VfL og formative vurderingsprosesser i kroppsøving (Georgakis & Wilson, 2012; Hay & Penney, 2013; López-Pastor m.fl., 2013), er det særlige bidrag fra denne avhandlingen forsøket på å studere VfL i en kroppsøvingsfaglig kontekst ut fra både elev- og lærerperspektiv ved samme skoler. Tilsvarende studier, med fokus kroppsøving der alle elever og lærere ved flere skoler inkluderes, er såvidt meg bekjent, ikke tidligere utført.

---

<sup>15</sup> Vurdering for læring heter «assessment for learning» på engelsk og er gjennomgående forkortet til AfL i engelske deler av avhandlingen.

Kappen har jeg valgt å skrive på norsk. Primært har det sammenheng med tre forhold. For det første et ønske om å kunne bruke språket mer nyansert enn hva jeg har erfart muligheter for når jeg benytter engelsk. Norsk er utvilsomt eneste språk hvor jeg føler meg hjemme, ikke minst med tanke på diskusjon av forskningsresultater. For det andre er det bare ved å skrive deler av avhandlingen på norsk at den eventuelt kan bidra direkte til norsk fagspråk og begrepsforståelse innenfor feltet vurdering i kroppsøving. Videre vurderte jeg det som meningsfullt å bruke norsk i lys av at det også var et mål med kappen å utdype og plassere artiklene inn i den norske utdanningskonteksten, hvorfra studienes empiri i utgangspunktet er innsamlet.

Jeg har tatt meg den frihet å holde en litt mer ordrik og personlig tone i kappen enn i artiklene. Fra (lesning og diskusjon av) vitenskapsteoretikere som for eksempel Lyotard (1988) og Feyerabend (1993) har jeg lært at resultater og tolkninger ikke blir mer vitenskapelig av at språket holdes i en form som indikerer objektivitet og avstand. Innimellom kan det være tvert om, at forskerens utvikling praktisk talt utgjør forskningens hovedfunn. Ikke at jeg mener det er slik i denne avhandlingen, men det uomgjengelige poeng er at ingen forskning kan presentere resultater der ikke mennesker på en eller annen måte står i relasjon til det utforskede.

Etter dette kapitlet, som har introdusert avhandlingens tema, bakgrunn og problemstillinger, følger i kapittel 2 en klargjøring av relevante begreper og de teoretiske horisonter avhandlingen skriver seg inn mot. Kapittel 3 plasserer avhandlingen vitenskapsteoretisk og redegjør for enkelte metodiske overveielser og valg. Kapittel 4 presenterer avhandlingens fire artikler med vekt på hovedfunn relatert til aktualiserte forskningsspørsmål. Kapittel 5 diskuterer studienes og resultatenes innbyrdes sammenheng og i hvilken grad den overordnede problemstilling er besvart, før Ronald får æren av å avrunde i en empirisk epilog som kapittel 6.

## Kapittel 2. Begreper, teori og kontekstualisering

Etter å ha fått mulighet til å lese (og diskutere) en god porsjon teori og forskningslitteratur om vurdering de siste fire fem årene, er det særlig to grunngivelse for formativ vurdering jeg er blitt hengende ved, eller som er blitt hengende ved meg. Den første er et argument som ble fremsatt allerede av Benjamin Bloom og hans kolleger ved *University of Chicago*. De hentet opp distinksjonen mellom formativ og summativ vurdering fra Scriven (1967) og overførte terminologien han hadde brukt i sammenheng med læreplanutvikling til vurderingsarbeid med egne studenter. Bloom og kollegene understreket læring som «a process which can be observed and evaluated as it is taking place» (Bloom, Hastings & Madaus, 1971, s. 138). På bakgrunn av dette argumenterte de for at studenter burde få vurderinger av eget arbeid og kunnskapsnivå mens de ennå hadde mulighet til å forbedre seg. Den klassiske avsluttende eksamen var dominerende ved universitetet, og den gav ikke studentene slike muligheter. Den var konstruert for å måle studentenes læring og kunnskapsnivå i etterkant av et undervisnings- eller opplæringsforløp, og vanligvis knyttet til rangering og seleksjon. I motsetning til denne type resultatvurderinger, som refereres til og betegnes summative, ble begrepet formativ vurdering knyttet til vurderinger som skulle informere, rettlede og justere, først undervisningen, men etter hvert også læreprosesser mens de ennå foregikk og studentene altså hadde forbedringsmuligheter. Den andre grunngivelsen jeg er blitt hengende ved, hevder at en art formativ vurdering alltid vil være til stede der læring skjer. Når noe skal læres, trengs det for eksempel informasjon om hva som skal læres, om hvilket utgangspunkt den som skal lære hva har, hva neste steg i læreprosessen vil være og eventuelt om det er behov for hjelp og støtte. Å svare på slike spørsmål står sentralt i det som presenteres som formativ vurdering, VfL inkludert, og viser hvordan vurdering forstått på denne måten innveves som en uomgjengelig del i læringsprosessene.

Stobart (2008) gjør et poeng ut av at VfL, i motsetning til mange teorier om læring, springer ut av empirisk forskning, og at dette taler for å ta særlig akt på det i vurderingspraksis og teoribygning. Det er således verd å merke seg at ingen av de to grunngivelsene jeg trakk fram er knyttet direkte til effektmål eller empirisk dokumenterte sammenhenger mellom læring og vurdering, hvilket utgjorde en vesentlig del av bakgrunnen da VfL ble løftet inn i internasjonal pedagogisk diskurs. I de par tiårene før Black og Wiliams (1998a, 1998b) nærmest epokegjørende studie som demonstrerte at vurdering, om den ble brukt riktig, holdt et betydelig potensiale til å øke elevenes læringsutbytte, hadde flere

reviewer vist at det var problemer med tradisjonell vurdering og karaktersetting (for eksempel Crooks, 1988; Natriello, 1987). Jeg kommer ikke til å forfølge det forskningsmessige og teoretiske belegget for VfL i dette kapitlet, men tydeliggjøre hvordan VfL og andre vurderingsbegreper<sup>16</sup> som validitet, reliabilitet, gjennomsiktighet, rettferdighet og sammenlignbarhet er benyttet og forstått i avhandlingen, samt kort skissere utviklingstrekk ved vurdering i kroppsøving i Norge og noen mulige utfordringer ved VfL.

### **Vurdering og vurderingens formål**

Vurdering brukes i skolesammenheng på flere måter. Om vurdering av skolen som helhet, vurdering av undervisning og vurdering av elever og deres kompetanse. Vurdering i denne avhandlingen handler om det siste, ofte også kalt elevvurdering. Begrepet elevvurdering kan føre assosiasjonene på ville veier, da det ikke er eleven, men elevens læring og faglige kompetanse som er siktemålet for vurderingen. Rent språklig har det også hersket uklarheter. Evaluering og vurdering har blitt benyttet om hverandre og omtrent synonymt i norsk skole. I tråd med det engelske skillet mellom «assessment» og «evaluation» ble det ved innføringen av LK06 bestemt at vurdering av elevers faglige kompetanse skulle hete vurdering. Tradisjonelt har det vært et skille mellom amerikansk og britisk engelsk på dette feltet, hvor karakterer og vurdering i britisk språkbruk har vært «grades» og «assessment», så har amerikanere vanligvis brukt «marks» og «evaluation». I avhandlingen følger jeg Kunnskapsløftet og bruker konsekvent begrepet vurdering om det som gjelder elevers faglige læring og kvalitet i læringsutbytte og kompetanse.

I Dobson, Eggen og Smith (2009) beskrives vurdering litt romslig som «...en gruppe prosesser som vi bruker når vi prøver å forstå og trekker slutninger om elevenes læringsprosesser, fremgang og læringsutbytte» (s. 24). Vurdering er da ikke en engangsaktivitet, men en langvarig prosess som inkluderer bruk av ulike vurderingsformer og verktøy. Det er viktig å forstå at vurdering aldri er en absolutt bedømming av elevers læring, men en fortolkning av elevers læring som konstrueres på bakgrunn av tilgjengelig informasjon.

Newton (2010) skriver at en av grunnene til mye forvirring når det kommer til vurdering er at vurdering er mye forskjellig og kan utføres med ulike hensikter, hvorpå han

---

<sup>16</sup> Flere av begrepene som brukes i vurderingsforskningen er begreper som også kan relateres til det metodiske i forskning (for eksempel validitet, reliabilitet og gjennomsiktighet). Jeg nevner dette fordi jeg underveis i avhandlingsarbeidet har erfart at dette felles begrepsapparat, som både brukes om vurdering knyttet til elevers faglige læring og kompetanse og for å omtale kvaliteten i forskning på elevers faglige læring og kompetanse, kan forvirre.



kategoriserer og presenterer 18 ulike formål med vurdering. Sammen med formative formål, hvor VfL tilhører, nevner han kategorier som seleksjon, sertifisering, veiledning, diagnostisering, monitorering av utdanningssystem, programevaluering, sammenligning og ressursfordeling. Et gjentakende problem er i følge Newton (2010) at vurdering utført for et formål feilkategoriseres. Det er likevel ikke nødvendig, hverken for forskere eller lærere, å forholde seg til alle formål vurdering kan ha. Black m.fl. (2003) opererer for eksempel med tre hovedformål – «accountability», sertifisering og læring – og bruker disse til å vise hvordan nasjonale og lokale krav til «accountability» og sertifisering påvirker vurderingsprosesser og implementering av VfL. Avhandlingens tredje artikkel (Leirhaug, MacPhail & Annerstedt, 2016) sier med referanse til Hay og Penney (2013) at de fleste vurderingsprosesser eller handlinger grovt kan sorteres enten for «accountability» eller for læring. Kroppsøvlingslærerne i Norge må håndtere begge disse. Først ved at LK06 og vurderingsregelverket er fulgt opp med en satsing på VfL og generelt krever arbeid med og dokumentasjon av formativ vurdering. Så ved at standpunkt karakteren i faget ved utgangen av grunnskolen, og ved endt videregående opplæring, er sluttvurderinger som har betydning for videre skolevalg og framtidige jobbmuligheter. Så selv om lærerens primære ansvar er elevens læring, noe som kan komme til uttrykk gjennom implementering av VfL og vurderingsarbeid på klasseromsnivå, understreker Hay og Penney (2013) med referanse til Popham (2009), at det tilhører den vurderingsliterate<sup>17</sup> lærer i settinger som den norske også å ha kjenskap til accountability-perspektivet<sup>18</sup>.

### **VfL i avhandling, teori og kroppsøving**

Leahy og Wiliam (2012) utpeker et kapittel av Harry Black (1986) med tittel *Assessment for learning* til å være første kjente bruk av VfL (på engelsk). Den kom altså omtrent samtidig som de nye perspektivene på sammenhenger mellom vurdering og læring begynte å bli tatt på alvor (Gardner, 2012; Wiliam, 2011). Innledningskapitlet til avhandlingen presenterte en overordnet definisjon hvor VfL plasseres i læringsituasjonen og forutsetter at lærere og elever er aktive deltakere som bidrar til at informasjon benyttes for å fremme læring. Ut fra denne

---

<sup>17</sup> For en redegjørelse av begrepet vurderingsliterate, se avsnittet «*Assessment literacy*» for kroppsøvlingslærere.

<sup>18</sup> Et hjemlig eksempel som aktualiserer dette er diskusjonen rundt PISA-undersøkelsene (se for eksempel Sjøberg, 2014). Dale (2010) viser til de tidlige PISA-resultatene fra århundreskiftet når han knytter utdanningsreformen Kunnskapsløftet til to fenomen: Det første er at læring og undervisning i skolen skal endres i samsvar med kunnskapssamfunnets krav til omstilling og livslang læring. Det andre er at resultatvurdering av elever skal vektlegges. I en mer journalistisk debatterende bok hevder Marsdal (2011) at den politiske understrømmen i kunnskapsløftet hele tiden har vært resultatkontroll. Elstad (2009) gir et godt bilde av hvordan media i denne sammenhengen kan virke inn mot skolene og deres pedagogiske praksis.

forståelsen og forskning på feedback er det formulert indikatorer for formative vurderingspraksiser. I sine seneste arbeider med teoriutvikling benytter Black og Wiliam (2009, 2012a) fem slike prinsipper eller nøkkelstrategier for VfL<sup>19</sup>:

- 1) Klargjøring og deling av kriterier for måloppnåelse
- 2) Å skape diskusjoner og situasjoner som kan avdekke læring
- 3) Framoverrettet feedback
- 4) Hverandrevurdering
- 5) Egenvurdering

De fem nøkkelstrategiene representerer en normativ kommunikasjon til lærere (og elever) om hva de bør tilstrebe i klasserommet for at vurdering skal støtte læreprosessen og fungere som et læringsfremmende verktøy. I avhandlingen benyttes disse for å konkretisere og operasjonalisere VfL i skolekontekst fra elev- og lærerperspektiv. Avhandlingens siste artikkel benytter de fem nøkkelstrategiene slik de er presentert. Første og andre artikkel anvender også VfL nøkkelstrategier, men da et sett på fire som MacPhail og Halbert (2010) formulerte med utgangspunkt i Black og Wiliam (1998b) til en studie av VfL i kroppsøving i Irland. Wiliam (2010) poengterer at oversikten over nøkkelstrategier ikke skal forstås uttømmende og at lærere er frie til å komme opp med andre strategier som også kan være læringsfremmende i den kontekst han eller hun arbeider innenfor. Med tanke på analysearbeid teoretisk innrammet av VfL nøkkelstrategier blir det derfor viktig samtidig å beholde en tilnærmet induktiv åpenhet for hva empirien kan komme opp med.

I VfL litteraturen rundt århundreskiftet hersket en forskningsbasert positiv optimisme over hva implementering av VfL kunne bety for elevenes læringsutbytte. Det var dette idealbildet av VfL og formativ vurdering som rundt om i verden motiverte skole- og vurderingsreformer med framtreddende VfL agendaer, Kunnskapsløftet inkludert. Ved tidspunktet som denne avhandlingsstudien ble planlagt, var Torrance (2007) en av få som

---

<sup>19</sup> Disse fem er hentet og oversatt fra et tabelloppsett opprinnelig foreslått av Leahy m.fl. (2005). Oppsettet er utviklet som en følge av at de krysset tre vurderings spørsmål (Where the learner is going? - Where the learner is right now? - How to get there?) med tre aktører (elev, medelev og lærer) som vanligvis er til stede og kan bidra til vurdering hva gjelder klasseromsvurdering og elevs læring. Oppsettet er gjengitt i tabellform i avhandlingens fjerde artikkel (Leirhaug & MacPhail, 2015). Punkt 4) og 5) er i oversettelsen forenklet ut fra Black og Wiliams (1998b) tidlige identifisering av VfL nøkkelstrategier, dette fordi det stemmer bedre med språkbruken i avhandlingsstudien datainnsamling og da spesielt spørsmål til elever gjennom spørreskjemaet. I engelsk original er disse «Activating students as instructional resources for one another» (4) og «Activating students as the owner of their own learning». Det er verd å merke at disse engelske beskrivelsene av strategier vil inkludere prosesser i klasserommet som jeg ikke uten videre vil beskrive eller klassifisere som vurderingsarbeid.

hadde hevet en kritisk røst mot VfL. Senere, i tiden parallelt med avhandlingsarbeidet, har det blitt satt spørsmålsteget både ved det empiriske grunnlaget som har ført til promoteringen av formativ vurdering, om faglig spesifitet er tilstrekkelig utredet og muligheter for genuin implementering av VfL i større skala (Birenbaum m.fl., 2015; Bennett, 2011; Torrance, 2012). Torrance (2007) advarte mot at tydeliggjøring av mål og kriterier kunne medføre en utilsiktet innsnevring av læreplanen, ved at elevene orienterte seg instrumentelt mot målene. Klargjøring av mål og intensiv veiledning og feedback knyttet til dette medfører en risiko for å fjerne

... the challenge of learning and reducing the quality and validity of outcomes achieved. This might be characterized as a move from assessment of learning, through the currently popular idea of assessment for learning, to assessment as learning, where assessment procedures and practices come completely to dominate the learning experience, and 'criteria compliance' comes to replace 'learning' (s. 282).

Det Torrance (2007) beskriver er at VfL kan føre til konsekvenser helt motsatt av det tilsktede. Til begrepsbruken må nevnes at «assessment as learning» gis en helt annen klang i denne innpakningen enn hvordan det samme ble presentert av Earl (2003) og benyttes av Gibbons og Kankkonen (2011) som ramme for å gjøre vurdering meningsfull for elever i kroppsøving. Vurdering som læring, som det heter på norsk, handler da om å hjelpe elevene til å forstå vurderingsprosessene og ta ansvar for egen læring. Det blir således nærmest en forutsetning for nøkkelstrategier knyttet til egenvurdering og har som mål å utvikle autonome og selvregulerte lærende elever. Dialog rundt lærings- og vurderingsprosessene vil kunne legge grunnlag for metakognisjon og demme opp mot den instrumentelle trussel som Torrance (2007) advarte mot. Black og Wiliam (2009) sier at dialogen også frambringer informasjon som vil kunne påvirke medelevenes evne til selv-regulering:

Self-regulated can be defined as multicomponent, multi-level, interactive self-steering process that target one's own cognitions, affects and action, as well as features of the environment of modulation in the service of one's goals (s. 13).

Selvstendige elever velger hensiktsmessige læringsstrategier og vurderer egen måloppnåelse. Dette forutsetter elever som til en viss grad er indre motivert for oppgaven, og overvåker og styrer læreprosessen deretter. Ikke alle elever er slik. Elevene i studien til James, Griffin og

Dodds (2009) uttrykte at vurdering var kjedelig. Dette er bare ett eksempel på utfordringer for kroppsøvingslæreren som skal iverksette VfL i egen undervisningspraksis. VfL nøkkelstrategier vil kunne arte seg ganske annerledes ovenfor den selv-regulerte og læringsorienterte elev sammenlignet med for eksempel elever som prøver å usynliggjøre seg selv i kroppsøving gjennom ulike skjuleteknikker (Lyngstad, Hagen & Aune, 2014). Elever som ikke er i stand til selv å regulere læringsprosessen avhenger av andre for å vurdere kvaliteten i arbeidet, hva som er neste oppgave og hva de trenger av støtte. I hvilken grad elevene er mottakelig for denne type hjelp, læringsorientering og faglig arbeid vil selvsagt variere og på bakgrunn av slike omstendigheter kom den første kritikken av formativ vurdering og VfL, båret fram blant annet av Perrenoud (1998) som hevdet at det egentlig ikke var et spørsmål om hvordan lykkes med læringsfremmende feedback, men at det heller gjaldt

...to concentrate on the theoretical models of learning and its regulation and their implementation. These constitute the real systems of thought and action, in which feedback is only one element (s. 86).

Som et svar på denne og lignende utfordringer har Black og Wiliam (2006, 2009, 2012a) i flere omganger drøftet sider ved VfL nøkkelstrategiene i lys av annen pedagogisk teori. Utgangspunktet er at VfL, og det som skjer i klasserommet, kan gis mening og plasseres i ulike teoretiske rammer. Å belyse og evaluere implementering av nøkkelstrategiene med ulike teoretiske innganger har ført til utvidete forståelser og reformuleringer. Når nøkkelstrategien egenvurdering reformuleres til å aktivere elevene som «owners of their own learning» (se note 19), så skjer det i dialog med forståelse av metakognisjon (Hacker, Dunlosky & Graesser, 1998), motivasjonsteori (Ryan & Deci, 2000) og attribusjon (Dweck, 2000). Black og Wiliam (2012a) utforsker nøkkelstrategier i kontekst av Engeströms (1987) kulturell-historisk aktivitetsteori og viser til hvordan klasserommet, eller læringsrommet, forstått som et «community of practice» (Lave & Wenger, 1991) eller en «figured world» (Holland m.fl., 1998) kan hjelpe med å bygge en forståelse av hvordan VfL nøkkelstrategier kan virke. Felles for disse teoriene er at de i et større bilde representerer en form for pedagogisk humanisme som «likestiller» lærer og elev som individer med egne meninger og interesser. Selv om teoritilnærmingen må kunne klassifiseres som eklektisk, demonstrerer den hvordan VfL og tilhørende nøkkelstrategier er godt forankret i et sosiokulturelt læringsperspektiv og slektskapet til Vygotsky og ideen om den proksimale utviklingszone er gjenkjennbar (Pryor & Crossouard, 2008).

Kritikken har vært viktig for å motarbeide oppfatninger av VfL som en magisk pedagogisk trylleformular eller «quick fix» for dårlig undervisning. Uten et læringsteoretisk rasjonale er det stor fare for at implementeringsforsøk og bruk av nøkkelstrategiene blir overfladiske og ikke fører til ønskede effekter og økt læringsutbytte. Det er sannsynligvis fundamentalt for å lykkes med iverksetting eller implementering av VfL nøkkelstrategier at det også skapes rom for refleksjon og teoretisk fordypning, for lærere kanskje helst i kontekst av kontinuerlig profesjonell utvikling. Ní Chróinín & Cosgraves (2013) mini innovasjonsstudie i kroppsøving med fem lærere indikerer at implementering av formativ vurdering og VfL strategier kombinert med (forsknings)samtaler kan føre til mer dyptgripende endringer i hvordan lærerne oppfatter kroppsøving og vurdering:

Teachers' perspectives on assessment changed dramatically as the value of assessment in enhancing the learning process was recognised. As a result of using assessment strategies, the teachers believed that the children learned more in their physical education classes and that they themselves became better teachers of physical education. (s. 230).

På den andre siden viser Tolgfors og Öhman (2015)<sup>20</sup> hvordan kroppsøvingslærere ikke trenger å oppleve denne omveltning i egen undervisning og tilnærming, men kan (oppfatte og) bruke VfL på andre måter og med andre formål enn læring, noe som igjen understreker behov for oppfølging og modererende refleksjon. Deres Foucault-inspirerte studie viser at lærere konstituerer ulike roller som igjen skaper ulike forventninger hos elevene. I tilfeller der elevene utsettes for en disiplinerende normaliseringsprosess i retning av det Torrance (2007) kalte «criteria compliance» er det per definisjon mulig å hevde at det ikke lengre er snakk om VfL, og at det i en vurderingsteoretisk sammenheng kan betraktes som et validitetsproblem. Validitet ved VfL handler overordnet om at utført vurdering må bidra til læring, ellers er det ikke valid<sup>21</sup>. At Tolgfors og Öhmans (2015) funn kan tolkes som et validitetsproblem løser ikke utfordringene de løfter fram, blant annet fordi vurderingens «andre formål» ikke er gjort eksplisitte og kan være usynlige for både elever og læreren selv. Det vil si at egenvurdering og hverandrevurdering gjennomføres, det gis tilbakemeldinger mot gitte mål, men det gjøres mot en bakgrunn av delvis skjulte hensikter, som disiplinering,

<sup>20</sup> Se også Tolgfors (2014). Et lignende perspektiv fra norsk videregående skole finnes i masteroppgaven til Palm (2014), selv om det er mer framtrødende at hans observasjoner og intervjuer viser få eksempler på bruk av læringsfremmende feedback og tanker som knytter vurdering til læring.

<sup>21</sup> Jamfør avsnittet *Validitet og VfL i kroppsøving*.

å organisere klassegruppen og opprettholde kontroll. Ved å følge Newton (2010), kan jeg si at aktørene feilkategoriserer vurderingens formål. For å få øye på, reflektere over og ta grep om slike nyanser trenger kroppsøvingslærerne å utvikle et felles – og kanskje fagdidaktisk tilpasset – begrepsapparat og språk. Et steg på denne veien, som i så fall sammenfaller med begreper som angår denne avhandlingens undersøkelse av implementering og forståelse av VfL i kroppsøving, kan være å se tradisjonelle vurderingsbegreper (som validitet) i lys av VfL. Etter et kort sveip gjennom norsk skolehistorie fram til vurdering i kroppsøving i LK06 – med vekt på videregående skole –, i den hensikt å kontekstualisere avhandlingsstudien, vil resten av dette kapitlet ta for seg noen av de mest vanlige vurderingsbegrepene før det avsluttes med å presentere en mulig ramme for å studere kroppsøvingslæreres kompetanse og kapasitet til å iverksette VfL som er valid, inkluderende og likeverdig.

### **Vurdering i kroppsøving, skolehistorisk skisse<sup>22</sup>**

Norsk skole har hatt en form for summativ vurdering rettet mot elevene i hele sin historie. I forkant av de første skolelovene der kroppsøving beskrives som eget fag (Synnestvedt, 1994), i 1848 for byfolkeskolen og med landsfolkeskoleloven av 1860, var vurdering i enkelte fag allerede formalisert for mange elever gjennom årlige eksamener og avgangsvitnemål. Karakteruttrykkene som ble benyttet varierte over hele landet, noe som gjorde det vanskelig å bruke dem til seleksjon og grunnlag for videre skolegang. Folkeskolelovene i 1936, som også gjorde kroppsøving til et obligatorisk fag for alle elever<sup>23</sup>, og Normalplanen av 1939 innebar en tydelig utviding av skolens oppgave til ikke bare å handle om fag, men også gjelde elevenes allmenndanning og personlige vekst. Det ble innført et felles bokstavkaraktersystem<sup>24</sup> for hele landet, og vurdering fikk et helt annet fokus enn i tidligere læreplaner.

I hele denne perioden, og mange steder et godt stykke inn i etterkrigstida, kan mye av vurderingspraksis i kroppsøving føres på kontoen for det López-Pastor m.fl. (2013)

---

<sup>22</sup> Ut over de kilder som er direkte referert i avsnittet er den sammenhengende framstillingen av utviklingstrekk og historiske detaljer basert på Augestad (2003), By (1998), Lysne (1999, 2004) og Skagen (1996).

<sup>23</sup> Dette er en vesentlig opplysning. For selv om kroppsøving var med i læreplanene var det først bare for gutter og opp til skolene om de ville tilby faget. Jacobsen (1966, s. 63) forteller at det i 1875 bare var 3,7 % av guttene som mottok undervisning i gymnastikk og militære øvinger. Om bildet endrer seg sterkt, spesielt i byskolene, og stadig flere skoler også starter gymnastikk for jenter i perioden fra århundreskiftet og fram mot 1930-åra, så signaliserer valgmuligheter om å tilby faget også at betydningen av vurdering i faget ikke er påaktet.

<sup>24</sup> Det felles bokstavkaraktersystemet var en femdelt gradering – Lite godt (Lg), Nokså godt (Ng), Godt (G), Meget godt (Mg), Særs godt (Sg) – og normalfordelingsprinsippet ble bestemmende for hvor stor prosent av elevene fra hvert årskull som skulle ha hver av karakterene. For grunnskolen bestod disse bokstavkarakterene til innføringa av L-97. Da ble systemet som var i bruk i videregående opplæring med tallkarakterer fra 1 til 6, der 1 er dårligst og 6 er best, gjort gjeldende for trinn 8-10 i grunnskolen.

internasjonalt beskriver som at «assessment, in so far as it existed, was straightforward» (s. 58). Med et innhold sterkt influert av militær eksersis, den tyske turnbevegelsen og svensk Ling-gymnastikk, etter hvert også for jenter og kvinner ut fra Elli Björkstens system, kunne læreren eller instruktøren stort sett observere direkte om eleven var i stand til å utføre de foreviste øvelser. Ut over dette handlet sannsynligvis vurdering mest om disiplin og det å følge kommando.

Idrett og ballspill fikk en alt mer sentral plass i kroppsøvingsfaget utover 1950-åra. Med merketaking og konkurranser hadde skoleidretten fått god utbredelse og skillet mellom kroppsøving og skoleidrett ble ved mange skoler utydelig. I *Læreplan for forsøk med 9-årig skole* (1960) av 1960 heter det sågar: «På høyere klassetrinn arbeider en med de øvinger som inngår i kravene til skoleidrettsmerket» (s. 23). Lek og ballspill var med, men det var idretten som først og fremst vant innpass i faget innpakket i medisinsk, teknologisk og helsemessig argumentasjon. Augestad (2003) skriver at stoppeklokke, målebånd og vekt ble «viktige instrumenter i medisinerens ønsker om å finne indikatorer på normalitet og progresjon. Men som nisse på lasset kom også rangeringen og konkurransene» (s. 194). Måletradisjonen og konkurranser kom nok også til å prege vurderingen i faget, understøttet av de idretters logikk hvor individers ulike fysiske prestasjoner rangeres etter objektive mål<sup>25</sup>. I den første midlertidige utgaven av *Mønsterplan for grunnskolen*, fra 1971, heter det under vurdering at en rekke prestasjoner i kroppsøving kan måles objektivt, og at læreren ut fra slike målinger kan danne seg et bilde av hvor eleven står på de felt målingen gjelder. I den endelige utgaven fra 1974 er dette tatt bort til fordel for at det er initiativ, spontanitet og fantasi som skal vurderes.

Endringen signaliserer et ønske om å dreie kroppsøving vekk fra konkurranser og at idrett udiskutabelt var sundt for alle. Faginternt kan ønsket om dreining knyttes til et tiltakende estetisk perspektiv og reformpedagogikkens inntog i kroppsøvingsfaget (Gurholt & Jenssen, 2007). Utad må det også tolkes i lys av den større politiske kamp og diskusjon av karakterer som forgikk i store deler av 1970-åra. Denne kampen førte blant annet til at norsk skole fra da av ikke har hatt vurdering med karakter før på ungdomstrinnene og oppover.

---

<sup>25</sup> Dette med bruk av fysiske tester av ulike slag ser ut til å være like omdiskutert som vedvarende når det kommer til vurdering i kroppsøving. Ennå i Prøitz og Borgens (2010) studie fra etter Kunnskapsløftet framtrer fysiske tester som noe det legges vekt på ved standpunktvurdering i kroppsøving. Flere masteroppgaver fra senere år tyder på at denne type testing er relativt utbredt. Græsholt (2011) fant at vurdering i kroppsøving sett fra elevenes perspektiv er mer preget av testkultur enn vurderingskultur. Risøy (2013), som kikket på testing spesielt, fant at det fortsatt benyttes tester som måler fysiske egenskaper som utholdenhet og styrke. Elevene uttrykker generelt lite læringsutbytte av de fysiske testene, og at testene først og fremst benyttes til karaktersetting og dokumentasjon. Rønninghaug (2011) hevder at hovedårsaken til at mange elever dropper ut av kroppsøvingsundervisningen er knyttet til tester og prestasjonskrav.

Fra 1976 ble kroppsøving gjort obligatorisk for elever ved studieretning for yrkesfag og fra da av har karaktervurdering i kroppsøving vært noe alle elever i videregående opplæring skal ha. Som del av arbeidet med å revidere fagplanen i kroppsøving for videregående opplæring ble det i 1983 utarbeidet retningslinjer i den hensikt å skape et grunnlag for en felles og mer forpliktende vurderingsordning. Disse presenterte en tredeling ved vurdering i kroppsøving hvor ferdigheter, innsats og kunnskap skulle vektlegges omtrent likt. Retningslinjene skulle «gjøres mer 'håndfaste' gjennom drøftinger rundt om på skolene» (Kårhus, 1992, s. 1). Selv om mange skoler og kroppsøvingslærere muligens opplevde at denne tredelingen fungerte bra, hevdet Blegenutvalget (NOU 1991:4, 1991) i 1991 at det eksisterte store forskjeller i «karakterkulturer» dels fordi standpunkt karakterer i stor grad måtte avgjøres ut fra faglig skjønn. Det ble pekt på «store ulikheter internt på den enkelte skole og skolene mellom» (s. 91). Kroppsøving var et av fagene med sterkt innslag av skjønnsmessig vurdering og Kårhus (1992) hevdet at sentrale aktører i kroppsøving og skolemiljøet fremdeles satt «med oppfatningen av at evaluering i faget varierer mye og at karaktersetting i kroppsøving fremdeles er en vanskelig sak» (s. 2).

Utfordringene med forskjeller i vurdering i kroppsøving gjaldt ikke bare videregående opplæring. Ellingsen, Nilsen og Oftedal (1985) beskriver hvordan kroppsøving i grunnskolen framstod ustrukturert og «kan sies å være et eksempel på hvordan Mønsterplanens funksjon som rammeplan fører til forskjeller innen et fags utforming og gjennomføring» (s. 1). Da karakteren i kroppsøving ved utgangen av grunnskolen fra 1983 ble gjort tellende ved inntak til videregående skole påpeker de et behov for at faginnhold og vurdering blir lagt til rette på en mer systematisk måte. Før dette hadde forskjellene i grunnskolen vært til å leve med fordi det var «utvist stor toleranse overfor utvalg av kriterier og krav til likhet» (s. 1).

Allerede ved karakterdebattene fra tidlig i 1970-åra hadde kritikken av normalfordelingsprinsippet nådd Norge. Alternativet som seilte opp var det mål- og kriteriebaserte systemet som Glaser og Klaus (1962) hadde presentert, og i Norge fikk vi det Lysne (2004) kaller en «tillempet målrelatert vurdering» der normalfordelingen fortsatt var styrende for eksamensresultater og en rettesnor for å studere standpunkt karakterer. Skolereformene i 1990-åra fastslo at vurdering av faglig kompetanse i Norge skulle mål- og kriteriebaseres, og med Reform 94 fikk videregående opplæring et forpliktende felles nasjonalt vurderingsgrunnlag. I følge *Læreplan for videregående opplæring: Kroppsøving* skulle det ved vurdering legges vekt på



- Kunnskap og evne til praktisk anvendelse av denne kunnskapen
- Ferdigheter
- Innsatsvilje og samarbeidsevne

Ved fastsetting av standpunktarakter i kroppsøving skulle disse telle 1/3 hver (KUF, 1994). Selv om det ved inngangen til 1990-åra ble gjort evalueringer av vurdering som også overveide alternative tilnæringer (Raaen, 1990), og det utvilsomt største utviklingsarbeidet med vurdering i kroppsøving i Norge ble gjort mot slutten av samme tiår (Hofnes 1997, 2000), er det liten tvil om at perspektivet på vurdering i kroppsøving nasjonalt fortsatt var sterkt preget av tradisjonelle spørsmål rundt karaktersetting. Forskingen som ble gjort (fortrinnsvis som hovedfagsoppgaver: Brattenborg, 1995; Hofnes, 1993; Ottesen, 1994; Peev, 2001) på vurdering i kroppsøving i Norge utover i 1990-åra, viser at det var spørsmålet om hva som ble vurdert, kriterier for karaktersetting og lærernes (ulike) vektlegging av disse som dominerte.

Folvik (1972) viser at problemer med tradisjonell karaktersetting i kroppsøving, og vurderingens mer progressive betydning i generell pedagogikk, var synlig i diskusjoner rundt kroppsøving i Norge allerede i 1960-åra<sup>26</sup>. Likevel kan jeg trygt slå fast at det først er med LK06 at formativ vurdering og VfL får en rolle å spille i faget. Det er likevel fortsatt vurdering med karakter som er i fokus når By m.fl. (2008) ut fra LK06 utarbeider en veiledning til vurdering i kroppsøving på videregående trinn 1, 2 og 3. Det kan også noteres at VfL ikke er nevnt i første revisjon av den kanskje mest brukte norske kroppsøvingdidaktikkboka (Brattenborg & Engebretsen, 2007)<sup>27</sup>. Det samme gjelder en svensk didaktikkbok som kom samme år, selv om den diskuterer formativ vurdering (Annerstedt, 2007). Da har det historiske sveip kommet til avhandlingsstudiens samtid og jeg vil i neste avsnitt bare gjengi avhandlingsrelevante hovedtrekk fra LK06 og tilhørende vurderingsregelverk.

### **Kroppsøving og vurdering i LK06**

Skolereformen Kunnskapsløftet gav kroppsøving for første gang en felles læreplan for hele det trettenårige skoleløpet. Alle læreplaner for fag har en lik struktur som presenterer fagets

---

<sup>26</sup> Folvik (1972) siterer en leder i bladet *Kroppsøving* (nr. 8, 1962) som bakgrunn for oppgavevalg: «Den form for karaktergivning en nå anvender stemmer ikke med synet som legges til grunn for undervisningen. Ferdighetskarakterer er et meningsløst mål på det vi har planlagt å gi elevene.»

<sup>27</sup> I revisjon av samme bok etter endringer i læreplan 2012, har LK06 begrep om underveisvurdering fått en styrket posisjon, og VfL er nevnt en gang (se Brattenborg & Engebretsen, 2013).

formål, hovedområder, timetall, grunnleggende ferdigheter, vurdering og kompetansemål. Læreplan i kroppsøving har kompetansemål som forteller hva eleven skal kunne etter 4. trinn, 7. trinn, 10. trinn og etter hvert av trinnene i videregående opplæring. På videregående trinn er faget inndelt i tre hovedområder: *idrett og dans*, *friluftsliv* og *trening og livsstil*. Innenfor hvert av disse beskrives kompetansemål. Totalt er det 25 kompetansemål for videregående trinn og her følger fire eksempler på hvordan de er formulert: Mål for opplæringa er at eleven skal kunne

- utvikle ferdigheter i individuelle idrettar og lagidrettar  
(Vg1, idrett og dans)
- skape dansekomposisjonar og vere med i dans som andre har laga  
(Vg1, idrett og dans)
- planleggje, gjennomføre og vurdere opphald i naturen, og gjere greie for korleis mat, utstyr og klede kan verke inn på opplevinga av naturen  
(Vg2, friluftsliv)
- praktisere og grunnngje trening som er relevant for å fremje god helse  
(Vg3, trening og livsstil)

Med hjemmel i Opplæringslova §13-10 (Opplæringslova, 1998) er skolene ved skoleeier pålagt å ha et forsvarlig system for vurdering. Forskrift til opplæringslova (2006) stadfester at alle elever har rett til vurdering: «Retten til vurdering inneber både ein rett til undervegsvurdering og sluttvurdering og ein rett til dokumentasjon av opplæringa» (§ 3-1). Sluttvurdering i kroppsøving får elevene som standpunktkarakter etter 10. trinn i grunnskolen og eventuelt på det tidspunkt de avslutter opplæring i faget i videregående opplæring. Dette betyr at all annen faglig vurdering per definisjon er undervegsvurdering, noe som også omfatter halvårsvurdering i form av karakter fra og med første halvår i 8. trinn. Karaktergradene er også beskrevet i forskriften med tall fra 1 til 6, der karakter 1 uttrykker at eleven har svært lav kompetanse i faget mens karakter 6 er uttrykk for fremragende kompetanse i faget.

For kroppsøving gjelder, som i alle andre fag, at grunnlaget for vurdering «er dei samla kompetansemåla i læreplanane for fag slik dei er fastsette i læreplanverket» (§ 3-3). I denne vurderinga «skal ikkje føresetnadene til den enkelte, fråvær, eller forhold knytt til ordenen og åtferda til eleven, trekkjast inn.»

Formålet med vurdering i fagene er «å fremje læring undervegs og uttrykkje kompetansen til eleven, lærlingen og lære kandidaten undervegs og ved avslutninga av opplæringa i faget» (§3-2). Om undervisvurdering heter det vidare at den «skal brukast som reiskap i læreprosessen, som grunnlag for tilpassa opplæring og bidra til at eleven aukar kompetansen sin i fag.» Dette forteller at all faglig vurdering fram til sluttvurdering har læringsstøttende og formative formål. Som et ledd i implementeringsarbeidet og til satsingen på VfL har Utdanningsdirektoratet (u.å.) formulert fire prinsipper som kjennetegn på god undervisvurdering. Forankret i forskning og lovverk sier disse at elevenes læring og forutsetninger for læring kan styrkes ved at elevene:

- 1) Forstår hva de skal lære og hva som er forventet av dem.
- 2) Får tilbakemeldinger som forteller dem om kvaliteten på arbeidet eller prestasjonen.
- 3) Får råd om hvordan de kan forbedre seg.
- 4) Er involvert i eget læringsarbeid ved blant annet å vurdere eget arbeid og utvikling.

Prinsippene kan gjenfinnes i vurderingsforskriften<sup>28</sup> og viser at VfL i LK06 har sammenfallende kjennetegn med, og bygger på samme forskning som, de prinsipper eller nøkkelstrategier som er lagt til grunn i avhandlingens metodetilnærming og forståelse av VfL.

Sammenfattet skal vurdering i kroppsøving i Norge fylle flere formål, hvorav VfL som del av undervisvurderingen vil være mest framtrædende i skolehverdagen. Ansvar for både dette kontinuerlige vurderingsarbeidet og å gi karakterer som danner grunnlag for opptak til videre utdanningsvalg og enkelte karriereveier tilfaller skolenivået, og i praksis ofte den enkelte kroppsøvingslærer. Videre, som det påpekes i en review fra OECD (2013), underkommuniseres gjerne betydningen av det spenningsforhold som ligger mellom forventningen om VfL og formative vurderinger i klasserommet og et nasjonalt vurderingssystem som også etterspør summative vurderinger som uttrykk for kvalitet.

### **Sammenlignbarhet og rettferdighet**

Sammenlignbarhet som vurderingsbegrep gjelder først og fremst summative vurderinger og angår i mindre grad formativ vurdering og VfL. Når det som i Norge kalles

---

<sup>28</sup> Prinsippene er forankret i § 3-1, § 3-11 og § 3-12 i Forskrift til opplæringslova (2006), der en finner omtrent likelydende krav og formuleringer knyttet til undervisvurdering.

standpunkt karakter eller sluttvurdering brukes til seleksjon for videre utdanning eller kvalifisering for yrke, forutsettes at denne betyr det samme i et fag enten den er fra en skole i Tana, Mo i Rana eller Flekkefjord. Om egenvurderingen i kroppsøving som blir gjennomført av en elev ved en skole er sammenlignbar med egenvurderingen gjennomført av en annen elev ved en annen skole er mindre interessant. Det vesentlig med egenvurderingen, her som et eksempel på en formativ vurderingsstrategi, er at den fungerer etter hensikten der den gjennomføres. Utfordringen ved VfL i forhold til sammenlignbarhet vil handle om at det er en sammenheng mellom praktisering og erfaring av VfL og summative vurderingsuttrykk, forenklet i LK06s terminologi til sammenheng mellom undervisvurdering og sluttvurdering. Gitt at egenvurdering og framoverrettet feedback fører til bedre læring i faget og at det er god sammenheng mellom slike undervisvurderinger og sluttvurdering, så vil det indirekte angå sammenlignbarhet at elever får like muligheter til formativ vurdering. Vinje (2008a, 2008b) gjør dette til et spørsmål om rettferdighet ved at mange kroppsøvingslærere rapporterer at de ikke makter å gi framoverrettet feedback til alle elevene. I Prøitz og Borgens (2010) studie av standpunktvurderingen ved to videregående skoler og fire ungdomsskoler kommer det klart fram at lærerne oppfatter sammenlignbarhet som en side ved rettferdighet:

En gjennomgående bekymring er hvor rettferdig egne standpunkt karakterer er i forhold til kollegaer man ikke samarbeider med om karaktergiving internt, og standpunkt karaktersetting eksternt som ved andre skoler i kommunen, fylke eller landet for øvrig (s. 89).

Rettferdighet er et begrep med lang historie og har vært knyttet til vurdering helt siden posisjoner i samfunnet skulle fortjenes ut fra andre prinsipper enn arvefølge. Allerede Aristoteles skrev at «rettferdighet» og «urettferdighet» ser ut til å bli brukt i mange betydninger<sup>29</sup>. Aristoteles fortsetter med at dette ikke er så lett å oppdage, fordi mange av betydningene ligner hverandre. Aristoteles poeng finner gjenklang hos Ronald, lærerstemmen som innledet denne avhandlingen, ved at han uttrykker å ha kommet til å forstå rettferdighet på en annen måte:

Ronald: I forhold til det som kan være vanskelig med vurdering, så er det veldig mye som kan være vanskelig [kort latter]. Men i hvert fall, sånn som min forståelse er blitt

---

<sup>29</sup> Aristoteles (1999, s. 86-87 i denne utgivelse/klassisk referanse: E N 1129a227)

etter hvert, så rettferdig hadde det kanskje vært noen som synes det var visst alt var likt. En målestokk å vurdere etter, noen mener det er rettferdig, ikke sant?

Forsker: Ja?

Ronald: Men sånn som jeg har skjönt det etter hvert, er vel at rettferdighet også kan være at elevene får en forespeiling av hva dette skal være og at det er klargjort i forkant, sånn at de har noe å forholde seg til. Om det er det ene eller det andre, det vil variere, og sånn er det lagt opp nå, at det varierer jo, fordi vi på skolen her lager en rettferdighet og noen andre lager en annen på en annen skole, så sånn sett er det jo at de er inneforstått med hva som skal skje som blir kunsten å få til.

Forsker: Å synliggjøre det?

Ronald: Ja, at det synliggjøres, og at det blir kjent og at det følges. At de føler at det blir tatt hensyn til, og at de får mulighet til å øve på det. Det er vel det som på en måte er det vanskelige.

Utdraget viser hvordan Ronald knytter rettferdig vurdering i kroppsøving til klargjøring av vurderingskriterier og vurderingsprosesser, mens for eksempel Vinje (2008b) og lærerne hos Prøitz og Borgen (2010) primært knytter det til forskjeller mellom skoler og sammenlignbarhet i vurderingsresultat. Disse to tilfellene er nok til å vise at «rettferdighet» er et begrep der konteksten blir avgjørende for innholdet i begrepet. Utdanningsforskningen peker på at rettferdighet i vurdering utfordres på ulike nivåer i skolesystemet. Rettferdig vurdering rører seg fra de store linjer som reproduksjon av sosiale forskjeller og overordnede politiske mål helt ned til det konkrete, som for eksempel at hvis alle elever skulle vurderes i høydehopp etter hvor høyt de hoppet, ville det bli et spørsmål om dette ikke var urettferdig i forhold til elevenes høyde. Davis (2009) argumenterer for at det er utfordringer med rettferdighetsbegrepet som ikke kan overvinnes gjennom forbedring av vurderingsprosesser. Disse problematiske sidene er mulig grunn til at for eksempel Miller m.fl. (2009) velger ikke å ta med rettferdighet som egen kategori når de skal diskutere kvalitet i vurdering. De velger heller å se rettferdighet i vurdering som en del av et rikholdig validitetsbegrep. På grunn av at denne avhandlingen primært er interessert i det som skjer på klasseromsnivå og erfaringer fra kroppsøving blir rettferdighet relatert til VfL behandlet i tråd med dette, selv om utfordringer med sosial reproduksjon (Dale, 2008) i skolesystemet vil gjelde for VfL også. Hay og Penney (2013) påpeker at (kritisk) oppmerksomhet mot mulige sosiale implikasjoner av vurderingspraksis vil bidra til det som er et nødvendig forhold mellom «the pursuit of a sosial just physical education and the optimisation of assessment validity» (s. 56).

### **Validitet og VfL i kroppsøving**

Validitet, eller det norske motsvarende gyldighet, er fundamentet for god vurdering. Validitet handler om at vurderinger vurderer det de skal vurdere. Det er et teoretisk hjelpemiddel for å avgjøre om hva det betyr å si at en informasjonstolkning, beslutning eller handling er bedre enn en annen. Validitet ble definert av Messick (1989) som «evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment» (s. 13). De fleste senere diskusjoner av vurderingsvaliditet i utdanningskontekster refererer tilbake til eller bygger på Messicks (1989, 1994, 1995) argumentasjon for at innholdet i, kriterier for og konsekvenser av vurdering i all utdanning kunne integreres i begrepet konstruktvaliditet<sup>30</sup>. Hovedtruslene mot validitet i fortolkning av elevprestasjoner og vurderingshandlinger er i følge Messick (1989) underrepresentasjon av konstruktet og konstruktirrelevante faktorer. Litt forenklet er konstruktet det som skal vurderes, men kan også peke på det som faktisk blir vurdert. Konstruktirrelevant er når en trekker inn andre faktorer eller elementer enn det som skal vurderes. Dette har vært et gjennomgående problem i kroppsøving<sup>31</sup>. Det er godt dokumentert, fra tidlige studier av Hensley m.fl. (1987) og Veal (1988), fram til senere studier både i USA (Young, 2011) og Sverige (Redelius, Fagrell & Larsson, 2009), at kroppsøvingslærere trekker inn elementer som holdning, innsats, bekleddning, oppmerksomhet og deltaking på tvers av standarder, læreplaner og vurderingsreguleringer som foreskriver at det er elevenes faglige kompetanse som skal utgjøre vurderingens grunnlag. Med intensjon om å studere VfL direkte fant Georgakis og Wilson (2012) utstrakt referering til tilsvarende konstruktirrelevante elementer blant australske idretts- og kroppsøvingslærere, mens de altså ikke fant eksempler på VfL praksiser<sup>32</sup>.

Underrepresentasjon av konstruktet betyr å vurdere en for snever del av det som skal vurderes. For å sette dette på spissen i en kroppsøvingskontekst vil sannsynligvis VO<sub>2</sub>max (kondisjonsmål basert på maksimalt oksygenopptak) ha betydning for langrennsprestasjon, men å måle bare VO<sub>2</sub>max for å vurdere langrennskompetanse gir lite mening. Dette er et

---

<sup>30</sup> Messick (1989, 1995) fremmet seks elementer som utgangspunkt for å gi en helhetlig validitetsvurdering (av et konstrukt) i utdanningskontekster; disse var «content», «substantive», «structural», «generalisability», «external» og «consequential». Jeg benytter ikke denne inndelingen i avhandlingen og har derfor ikke utdypet dem videre.

<sup>31</sup> Med tjue års mellomrom diskuterer både Carrol (1994) og Hay og Penney (2013) dette karaktertrekket ved funn fra forskning på vurdering i kroppsøving.

<sup>32</sup> Dette refereres i kappens innledende kapittel og utgjør en viktig horisont for avhandlingens fjerde artikkel (Leirhaug & MacPhail, 2015).

tildratt og ekstremt eksempel på underrepresentasjon. Mer direkte knyttet til dagens læreplan i kroppsøving (LK06) vil kunne være at fair play ikke vektlegges ved vurdering i kroppsøving, selv om det understrekes i formålet for faget og det er egne kompetansemålformuleringer på ulike trinn i læreplanen som forteller at fair play er noe elevene skal praktisere og lære.

Forskningen på validitet ved vurdering i kroppsøving internasjonalt har i senere år i stor grad benyttet «alignment», sammenheng mellom eller innretning av, Bernsteins (1975) «three message systems for schooling» – «pedagogy», «curriculum» og «evaluation»<sup>33</sup> – som utgangspunkt for å drøfte valide vurderingskriterier og kvalitet i kroppsøving<sup>34</sup>. Dette er ikke uten videre uproblematisk når det kommer til formativ vurdering og VfL. Jeg minner om at Bernstein forsket og skrev fram denne teorien omtrent samtidig med at begrepet formativ vurdering ble tatt i bruk. Systemet vurdering i hans «message systems» handler derfor rimeligvis om en praksis dominert av summativ vurdering. VfL foregår underveis i læreprosessen og lykkes kanskje aller best inkorporert eller som integrert del i undervisning og læringsarbeid. Dette betyr at VfL plassert i Bernsteins terminologi i like stor grad vil tilhøre systemet «pedagogy» som «assessment», og kan altså ikke forstås som sammenfallende med systemet vurdering. Denne overlappen, som følge av nyere vurderingsperspektiver, kan være en av grunnene at Hay og Penney (2013) legger ekstra stor vekt på Bernsteins egen påpekning av systemenes interne sammenheng. Denne overlappen er også grunnen til at avhandlingen ikke forstår vurdering som et «message system» og ikke legger «alignment» som en nødvendig forutsetning for validitet, men inkluderer det som ett av flere mulige validitetskjenne tegn ved VfL i kroppsøving<sup>35</sup>.

VfL kan ha både tilsiktede og ikke-tilsiktede effekter på elevens læring og mål, som all annen vurdering, kunne demonstrere validitet. Det avgjørende elementet for VfL er i følge

---

<sup>33</sup> «Evaluation» er i tråd med datidens språkbruk og en praksis som langt på vei er opprettholdt i amerikansk utdanningskontekst, og således kan det forstås tilsvarende det engelske «assessment» og vurdering i LK06 og denne avhandlingen. Bernstein (1975) beskriver sammenhengen mellom de tre systemene slik: «Curriculum defines what counts as valid knowledge, pedagogy defines what counts as valid transmission of knowledge, and evaluation defines what counts as a valid realization of this knowledge on the part of the taught» (s. 85).

<sup>34</sup> Se for eksempel Hay og Macdonald (2008), Hay og Penney (2009, 2013) eller Redelius og Hay (2012). Arnesen, Nilsen og Leirhaug (2013) er et eksempel på bruk av de «three message systems» læreplan, undervisning og vurdering i norsk kroppsøvingkontekst. Jeg tror ikke det er urimelig å hevde at den utstrakte bruken av Bernsteins (1975) «message systems» kan knyttes til Peter Hay og hans formidable bidrag til forskningslitteraturen på forståelse og konsekvenser av vurdering i kroppsøving de siste ti år.

<sup>35</sup> Understreker at dette ikke tilsidesetter «alignment» mellom de tre «message systems» som et godt rammeverk for å styrke læreres planlegging og forståelse av vurdering i kroppsøving som ledd i å bedre kvaliteten i kroppsøving. Dette gjelder også for eksempel Georgakis og Wilsons (2012) tilsvarende drøfting på bakgrunn av Biggs (1996) «constructive alignment», eller det Klenowski og Wyatt-Smith (2014) beskriver som «front-ending assessment» – undervisningsplanlegging som i praksis starter med spørsmål om hvordan informasjon om elevens læring skal frembringes.

Stobart (2012) om vurderingsarbeidet oppfyller formålet om å fremme læring. I så måte er det et konsekvensielt validitetsargument. Dette gjør det vanskelig, for ikke å si – i prinsippet – umulig, på forhånd (i planlegging) å avgjøre hva som vil være konstruktirrelevant. Selv om forskning på feedback (for eksempel Hattie & Timperley, 2007) kan gi en pekepinn på hvilke type tilbakemeldinger som for flest elever vil føre til videre faglig læring, og selv om VfL implementeres og gjennomføres med gode intensjoner, er det ikke sikkert at denne praksis faktisk fører til økt læring. Med dette som utgangspunkt oppsummerer Stobart (2012) validitet i VfL og andre formative vurderingspraksiser slik:

Validity is then about establishing what has contributed to further learning and what has got in the way of this. This will include the cultural and learning context, the quality of classroom interactions, the teacher's and learner's clarity about what is being learned and the effectiveness of feedback. (s. 241).

Validitet knyttes ut fra dette i avhandlingen tett til hvordan VfL «pakkes ut» eller operasjonaliseres som nøkkelstrategier. Den del av validitetsvurderingen avhandlingens studier kan besvare, er i hvilken grad hver av nøkkelstrategiene, og andre aspekter ved vurderingspraksis, er implementert og forstått på måter som kan fremme eller hemme elevers læring. Ideelt skulle jeg her kunne skrevet i hvilken grad hver VfL nøkkelstrategi «har bidratt til elevenes læring», men metodedesignet er ikke utformet for å kartlegge elevenes læringsutbytte på den måten.

Selv om konsekvensiell validitet er avgjørende ved VfL, er konstruktvaliditeten totalt sett fortsatt avhengig av opplærings- og læreplanmål, faglig innhold og hvilke kriterier kroppsøvingslærere vektlegger ved vurdering. Redelius (2007) argumenterer for at spørsmål om hva som skal vurderes og hvordan det vurderes kan ha pedagogiske konsekvenser og dypere betydning for legitimering og oppfatning av kroppsøving. Endring i vurderingsordning kan føre til endring i prioritert fagstoff og hvilke typer kunnskap som verdsettes. Parallelt med dette viser Melograno (2007) at hvilke kriterier som legges til grunn ved vurdering i kroppsøving kan ha store konsekvenser for hvilket resultat (i karakterer) elevene kommer ut med. Hans mest ekstreme eksempel er en elev som starter med praktisk talt topp mestring og kompetanse i faget, og derfor går gjennom opplæringen uten at kompetansen er synlig forbedret. Om mestring og kompetanse vektlegges vil denne eleven få toppresultat, mens en ren vektlegging av framgang vil resultere i stryk for den samme eleven. Perspektivet innebærer også at elever med betraktelig dårligere mestring og kompetanse



sammenlignet med andre elever likevel vil kunne få bedre karakter i faget. Det er flere momenter i Melogranos (2007) eksempel som kan diskuteres, blant annet forståelsen av framgang. Her var imidlertid hensikten bare å vise betydningen av at vurderingskonstruktet framstår veldefinert og bygger på åpenhet og felles forståelse blant lærere og elever. Effektiv framoverrettet feedback og andre VfL nøkkelstrategier er effektive bare så lenge de orienteres mot det som skal læres eller gjennomføres. Om det framkommer store forskjeller mellom lærere i bruk av kriterier og hva de vektlegger, så indikerer det utfordringer for validiteten ved vurdering i faget. Om lærere som forholder seg til samme læreplanmål og felles kriterier likevel kommer til ulik karakter (eller annen summativ vurdering), er det et tilfelle som tradisjonelt er å betrakte som et reliabilitetsproblem. Er det slik også ved VfL?

### **Reliabilitet og gjennomsiktighet**

Reliabilitet, også omtalt som pålitelighet, opptrer ofte side om side med validitet. Litt folkelig sagt handler reliabilitet om at det sluppmessige får minst mulig påvirkning på resultatet. Det handler om i hvilken grad en vurdering og dens konsekvenser vil være de samme om vurderingen gjøres i dag eller om et år, og i hvilken grad en elevs læring og kompetanse vil vurderes likt av forskjellige lærere. En elev kan også prestere bedre den ene dagen enn den andre, noe som vil kunne ha stor betydning for reliabiliteten i vurderinger basert på enkelttilfeller. Motsatt er det argumentert for at læreres karaktersetting i utgangspunktet må betraktes å ha høy reliabilitet fordi de baserer seg på en rekke ulike vurderinger gjort over lang tid, og gjerne i dialog med den lærende.

Forholdet mellom reliabilitet og validitet kan beskrives som at lav reliabilitet i en vurdering vil undergrave validiteten, mens høy reliabilitet vil være praktisk talt verdiløs hvis validiteten ikke er tilfredsstillende. Ved formativ vurdering, VfL inkludert, vil lav reliabilitet i klassisk forstand ikke nødvendigvis føre til utfordring ved validiteten. To lærere kan tolke samme vurderingsinformasjon ulikt og gi eleven framoverrettet feedback på forskjellige måter. Hvis utfallet i begge tilfeller er økt læringsutbytte, så er det overordnede validitetskriteriet for VfL oppfylt. Black og Wiliam (2012b) tar utgangspunkt i sin egen definisjon (se note 7), og peker på at formative vurderinger, VfL inkludert, er reliable i den grad

...that the assessment processes being used generate evidence that consistently lead to better, or better founded decisions, which turns out to be a less restrictive condition than is needed for summative assessment (Black & Wiliam, 2012b, s. 260).

Hay og Macdonald (2008) skriver at når standardiseringsgraden i vurderingene går ned, noe som altså er tilfelle ved VfL, blir begrepene reliabilitet og validitet vanskeligere å skille. I kroppsøving gjelder dette ikke bare formative vurderingspraksiser. For så lenge kroppsøvingslærere refererer til en form for «magefølelse» eller intuisjon for å grunngi karaktersetning kan standardiseringsgraden i praksis neppe kalles høy. Dette er muligens noe av bakgrunnen for at for eksempel Annerstedt og Larsson (2010) ikke adresser begrepene separat, men omtaler reliabilitet og validitet som om de var ett.

VfL i et skolesystem med metodefrihet og desentralisert ansvar for vurdering bygger på at skolekulturer og undervisning, så vel som elever, er forskjellige. Det innebærer at vellykket VfL ved en skole og i en kroppsøvingssetting ikke nødvendigvis vil fungere i en annen. Dette åpner for at forskningens påvisning av at kroppsøvingslærerne praktiserer vurdering ulikt også kan borge for kvalitet. Resultatene til Prøitz og Borgen (2010) tyder på at lærerne selv forstår rettferdighet i retning av like kriterier. Dette kan muligens koples til Engvik (2013), hvor kroppsøving utmerker seg som det faget der flest lærere etterlyser nasjonale føringer for vurderingen. Ronalds versjon hvor «vi på skolen her lager en rettferdighet og noen andre lager en annen på en annen skole» henter mot en annen tilnærming der ulik vurderingspraksis i seg selv er ikke problemet.

Forskningen på vurdering i kroppsøving og tilhørende teoribygning virker så langt i liten grad å ha tatt opp i seg at det ikke er noen automatikk i at like kriterier og lik vurderingspraksis fører til høyere validitet. Men da kan forskjellene mellom lærernes grunnvinger selvsagt ikke skyldes at lærerne har vanskeligheter med å verbalisere egne kriterier og viser til en form for «magefølelse» som utgangspunkt for vurderinger (Annerstedt & Larsson, 2010; Hay & Macdonald, 2008). Gjennomsiktighet, også omtalt som transparens, forutsetter at kroppsøvingslærerne har god forståelse av egen vurderingspraksis og utgjør et fundament for i det hele tatt å kunne diskutere en sprikende vurderingspraksis som uttrykk for validitet og likeverd. Gjennomsiktighet som vurderingsbegrep overlapper deler av VfL nøkkelstrategier ved at det handler om å synliggjøre vurderingskriterier og prosesser. Videre er det en forutsetning for å utvikle egen- og medelevvurdering, så vel som selv-regulering. Det handler om en form for avprivatisering som er nødvendig for å utvikle en felles vurderingskultur ved skolen, og Engvik (2013) hevder at kroppsøvingslærere som arbeider innenfor et vurderings- og tolkningsfellesskap utvikler egen vurderingskompetanse og har bedre forståelse av vurderingspraksis.

### «Assessment literacy» for kroppsøvingslærere

Avhandlingens tredje artikkel (Leirhaug, MacPhail & Annerstedt, 2016) benytter et teoretisk rammeverk for «assessment literacy» i kroppsøving som er foreslått og utviklet av Hay og Penney (2013). I artikkelen trekkes en kjapp historisk linje som forteller at «literacy» opprinnelig er et begrep utviklet i tilknytning til tekstforståelse og evne til å lese og skrive. Jeg viser til artikkelen for en beskrivelse av hvordan «literacy» har blitt vanlig på mange felt<sup>36</sup> og mer eksakt hvordan Hay og Penney (2013) bruker en utvidet forståelse av begrepet hvor «the capacity to actively participate in a social field is fundamental to what it means to be literate» (s. 71).

Literacy er et begrep som ikke til vanlig er knyttet til vurdering i norsk språkbruk. Et enkelt søk (i 2015) med søkemonitoren til Google gav ett meningsfullt treff for uttrykket «vurdering literacy», men da på en dansk nettside. Søk på «vurderingsliteracy» og det fornorskede «vurderingsliterate» ga ingen treff. Det er mer vanlig å snakke om vurderingskompetanse, og det er ikke urimelig å hevde at kompetansebegrepet som ligger til grunn i LK06 er nært beslektet med det engelske «literacy». Men det er nettopp et poeng for Hay og Penney (2013) at «assessment literacy» handler om mer enn kunnskap og vurderingskompetanse. Ved å innføre begrepet kapasitet ved siden av kompetanse blir det avgjørende både i hvilken grad aktørene er i stand til delta effektivt, medskapende og kritisk i den (til enhver tid) samtidige fagdidaktiske vurderingsdiskurs, og dermed hvilke vurderinger som faktisk gjennomføres. I den norskspråklige delen av avhandlingen buker jeg innimellom den norske varianten vurderingsliterate om lærere og elever som er i stand til å delta i og bidra til å utvikle vurderingsdiskursen, og dette brukes da synonymt med de steder jeg referer til det engelske «assessment literacy».

Hay og Penney (2013) presenterer fire elementer<sup>37</sup> med en indre sammenheng for å beskrive hva det å være vurderingsliterate innebærer. Den vurderingsliterate kroppsøvingslærer er kjent med læreplan og vurderingsreguleringer, kan lage og kommunisere vurderingskriterier og, ikke bare har kunnskap til, men er også ut fra rammer og tilgjengelig ressurser i stand til å iverksette gjennomtenkte vurderingspraksiser. Videre må læreren kunne fortolke innhentet informasjon meningsfullt og delta i en kontinuerlig dialog rundt vurderingspraksis for å motvirke sosial urettferdighet og potensielt negative

---

<sup>36</sup> I tillegg til eksemplene som nevnes i artikkelen passer det å vise til at «physical literacy», slik begrepet er utviklet av Whitehead (2010), nylig er presentert for første gang på norsk og relatert til norsk kroppsøvingskontekst i boka *Fysisk aktiv læring* (Vingdal, 2014).

<sup>37</sup> De fire er «assessment comprehension», «assessment application», «assessment interpretation» og «critical engagement with assessment». Se Leirhaug, MacPhail og Annerstedt (2016) og presentasjon av artikkelen i kappens kapittel 4.

konsekvenser for enkeltelever. Den vurderingsliterate kroppsøvingslærer i Norge anno 2011-2012 må følgelig – ut fra LK06 og tilhørende VfL-agenda – forventes å ha iverksatt VfL som del av sin praksis, eller i det minste ha kunnskap om VfL og være i stand til da å forklare eller kritisk reflektere rundt hvorfor dette ikke er implementert og lar seg gjøre.

Selv om «assessment literacy» knyttet til formative vurderingspraksiser og VfL primært vil gjelde klasseromsnivået og det som skjer i arbeidet med elevene, forventes den vurderingsliterate kroppsøvingslærer også å ha forståelse for hvordan det Popham (2009) kaller «accountability assessments» og selve skolesystemet kan påvirke vurderingsarbeidet og læreres valg på klasseromsnivået. Dette betyr ikke at kroppsøvingslærerne må ha inngående kunnskap om alle mulige typer vurdering for å lykkes med gode formative vurderingspraksiser. Threlfall (2005) er en av dem som uttrykker dette presist når han skriver at lærere «do not need to have developed pedagogical knowledge for every assessment that could be done, as they will only be doing assessments they have planned for, for which they do have an adequate level of expertise» (s. 62).

Hay og Penney (2013) mener de gjennom utlegningen av «assessment literacy» for lærere og elever i kroppsøving har bidratt med et begrepsapparat som kan hjelpe til med å få øye på, drøfte og utvikle den kunnskap og kompetanse som er nødvendig for å gjøre vurderingsarbeidet i kroppsøving (mer) effektivt, valid og sosialt rettferdig. På samme tid understreker de at «assessment literacy in physical education is a new and undeveloped idea» (s. 81). Det ligger ingen egen empirisk undersøkelse til grunn for deres utlegning og promotering av «assessment literacy». De bygger teoretisk på tidligere vurderingsforskning i kroppsøving og skriver at for å komme videre «is the need for empirical work focusing on teachers' existing assessment literacies» (s. 81). Så da passer det godt å ta steget over i neste kapittel som omhandler og kommenterer avhandlingens metodologi.

### **Kapittel 3. Metodologiske detaljer, overveielser og valg**

Kapitlet grunngir forskningsmessige og metodiske valg som angår avhandlingen som helhet, posisjonerer tilnærmingen vitenskapsteoretisk og utdyper noen metodiske detaljer. I tillegg presenteres forskningsetiske betraktninger og handlingsvalg.

#### **Forskningsstrategi og design**

Innledningen til denne avhandlingen presenterer flere av de forskningsstrategiske viktige valgene: Hvorfor VfL? Hvorfor videregående skole? Videre at den norske konteksten etter innføringen av LK06 gir mulighet til å studere implementering av VfL som del av en læreplan- og vurderingsreform på systemnivå, og at den type studier ikke er gjort med kroppsøving i hovedfokus. Det var forventet at utfordringer og spørsmål om implementering, så vel som forståelse, av VfL kunne arte seg annerledes når en skalere opp fra studier på klasserom- og lærernivå til å gjelde hele skoler som del av et større utdanningssystem (Leahy & Wiliam, 2012; Thompson & Wiliam, 2008; Torrance, 2012). Innledningsvis ble det også vist til at lite av vurderingsforskningen i kroppsøving har perspektiv fra både elever og lærere. I lys av dette ble det besluttet å gjøre utvalg på skolenivå, der en inkluderte studier både med elev- og lærerperspektiv ved samme skole for å studere implementering og forståelse av VfL i kroppsøving som et resultat av LK06 og det fornyede vurderingsfokus. Etter å ha tatt prosjektets rammer i betraktning førte overveielsene til et valg om å studere VfL i kroppsøving ved seks skoler; nærmere beskrivelse av utvalg og kriterier følger nedenfor.

Med skolenivået som utvalgsenhet ble en kvantitativ tilnærming gjennom spørreskjema valgt for å kunne representere VfL i elevperspektiv ved alle seks – og hver av – skolene som en helhet<sup>38</sup>. Ut fra en vurdering av antallet i populasjonen kroppsøvingslærere og med mål om å danne dypere og mer utfyllende drøftinger av implementering og forståelse av VfL i kroppsøving ved skolene, ble en kvalitativ tilnærming gjennom fokusgruppeintervju valgt for å innhente lærernes perspektiver. Fokusgruppeintervju ble funnet hensiktsmessig fordi det var skolenivået og kroppsøvingslærernes forståelse og beskrivelser av vurderingspraksis ved den enkelte skole, og ikke den individuelle lærer, som var studiens

---

<sup>38</sup> Det ble også vurdert kvalitative tilnærminger – som for eksempel intervju eller observasjon – for å utdype elevperspektivet, men det ble ikke fulgt opp på grunn av prosjektets rammer.

hovedinteresse. Med intensjon om å kunne gjøre sammenligninger skolene mellom, ble kroppsøvlingslærerne ved hver enkelt skole samlet i en fokusgruppe.

Bruk av kvantitativ og kvalitativ metode for innsamling og prosessering av empiri i avhandlingsstudien gjør at den som helhet kan beskrives som et *mixed method design* (Creswell, 2009; Tashakkori & Teddlie, 2003). I Tashakkori og Teddlie (2003) drøftes ulike former for metodeblanding, under hvilke forutsetninger metoder kan kombineres og hvordan dette kan medføre særlige utfordringer forbundet med kvalitetsvurdering. I denne avhandlingen betraktes metodetilnærmingene som likestilte kunnskapsprodusenter med tanke på den overordnede problemstillingen, men enkeltstudiene tatt i betraktning så skaper den faktiske metodeblanding få metodologiske utfordringer. Bruken av kvalitativ og kvantitativ metode for innsamling av data er rettet mot to ulike populasjoner og det er egentlig bare i avhandlingens første artikkel (Leirhaug & Annerstedt, 2015) at kvalitativt og kvantitativ materiale analyseres mot hverandre. De tre andre artiklene benytter enten elevdata fra spørreundersøkelsen, og representerer da en tradisjonell kvantitativ tilnærming, eller lærerdata fra fokusgruppene, og da med tilhørende anerkjente kvalitative analysetilnærminger.

Ut fra Bryman (2012) og Ringdal (2013), som begge opererer med fem hovedtyper<sup>39</sup> forskningsdesign, kan avhandlingens studie også beskrives som et tverrsnittdesign. I en tverrsnittstudie innsamles data bare en gang for hver analyseenhet eller informant, og da vanligvis innenfor et gitt avgrenset tidsrom. Dette er tilfellet for både spørreundersøkelsen som gir elevdata og fokusgruppeintervjuene med kroppsøvlingslærere som gir lærerdata til avhandlingens enkeltstudier. Hensikten er å beskrive og fortolke en nåsituasjon, eller status til, vurdering i kroppsøving med vekt på implementering og forståelse av VFL, der utgangspunktet er empiri samlet ved seks videregående skoler.

### **En pragmatisk og sosiokulturell informert virkelighetsforståelse**

Kvantitativ og kvalitativ metodetilnærming har opphav i det som gjerne skisseres som ulike vitenskapsteoretiske kunnskapssyn (Bryman, 2012; Tashakkori & Teddlie, 2003). Sterkt forenklet knyttes da utviklingen av kvantitativ metode til mer positivistiske vitenskapsideal, hvor det antas at det finnes en sann(ere) kunnskap om fenomener i verden og at vitenskapens

---

<sup>39</sup> Hos Ringdal (2013) heter disse eksperimentell design, tverrsnittdesign, langsgående design, casedesign og komparativ design. Det presiseres at ulike design kan kombineres og at det finnes alternativer til en slik inndeling, blant annet i forhold til *mixed method* som på norsk da omtales som *flermetodisk design*. Ringdal (2013) klassifiserer både bruk av kvantitative spørreundersøkelser og kvalitative samtaleintervju i et lite utval personer som «svært vanlig» ved tverrsnittstudier (s. 106).

oppgave er å jakte på og avdekke denne. Kvalitativ metode kan på sin side knyttes til et hovedsyn der kunnskap er noe som konstrueres og gis legitimitet i og gjennom sosiale, politiske og språklige kontekster, og vitenskapens hovedoppgave er å bidra til en utdypet forståelse av fenomener og hvordan de konstrueres som fenomener i verden. Ringdal (2013) og Creswell (2009) skriver at senere utviklingstrekk i metodevalgdiskusjoner har gjort det mer vanlig å se kvalitative og kvantitative tilnærminger som komplementære snarere enn som motsatser. Uavhengig av metodenes underliggende vitenskapsteoretiske og historiske ballast vil det forskningen spør etter, problemstillingene, og forskernes kompetanse utgjøre fundamentet for kvalitativt gode metodevalg. Creswell (2009) argumenterer videre, i denne sammenheng fra en *mixed method* posisjon, for at forskere bør tilstrebe å gjøre eksplisitt det verdensbilde eller de filosofiske virkelighetsforståelser som ligger til grunn for et forskningsprosjekt. Dette vil bidra til å forklare hvorfor en metodisk tilnærming er å foretrekke foran en annen. Creswell (2009) bruker det engelske begrepet «worldview» for å betegne et slikt mer eller mindre artikulert grunnleggende sett av oppfatninger som vil farge erfaringer og bidra til å styre handlingsvalg. Andre snakker om «paradigms» (Mertens, 1998), eller «ontologies» og «epistemologies» (Bryman, 2012). Uten å gå inn på meningsdifferanser begrepene mellom holder jeg meg til virkelighetsforståelse.

Jeg sier i avsnittets overskrift at avhandlingen har en pragmatisk og sosiokulturelt informert virkelighetsforståelse. Med informert ønsker jeg å signalisere at det tilhører avhandlingens virkelighetsforståelse at enhver beskrivelse av, eller forsøk på å artikulere, en virkelighetsforståelse må uttrykkes fra et ståsted eller perspektiv, og således alltid allerede selv innebærer å være uttrykk for en virkelighetsforståelse. Denne grunnleggende utfordringen, som er et eksempel på det filosofien kaller en uendelig regress, taler etter min mening for at den ekspliseringen av virkelighetsforståelser som Creswell (2009) og andre etterspør bør ha en grad av ydmyk fleksibilitet. Derfor «informert».

En sosiokulturell inngang baseres på en forståelse av at opplæring og skole foregår på en sosial arena og innebærer interaksjon mellom mennesker, samtidig med at det forteller noe grunnleggende om kunnskapssynet som legges til grunn for å forstå kroppsøving og VfL. Både internasjonalt (Evans, Davies & Penney, 1999; Green, 2008; Kirk, 2010) og i senere års norsk kroppsøvingsforskning presenteres faget kroppsøving som en sosiokulturell konstruksjon<sup>40</sup> (Moen, 2011; Svendby, 2013; Lyngstad, 2013). I korte trekk innebærer et

---

<sup>40</sup> En slik oppfatning av hva og hvordan et skolefag oppstår og formes er generell og gjelder ikke kroppsøving spesielt. For eksempel skrev Ludvigsen-utvalget i deres underveisrapport om fremtidens skole at skolefagene

sosiokulturelt perspektiv at kroppsøving ikke er noe gitt, men vil være preget av land og kultur, historie og samfunnsutvikling, læreplaner og etablerte praksiser, og at faget og dets legitimering stadig vil være i en omskappingsprosess og endres over tid. Det er i lys av en slik forståelse at Tinning (2012), med et memetisk perspektiv på kroppsøvingsfagets framtid, kan erklære at uansett hva idéen eller hensikten med kroppsøvingsfaget vil være, så vil det være en som er i takt med tiden, eller som han skriver, «'fit' for the contemporary environment» (s. 234). Men en sosiokulturelt informert virkelighetsforståelse inviterer også til å søke etter alternativer, i kraft av at dominerende forståelser og praksiser, det bestående, er skapt gjennom prosesser i et sosialt, politisk og kulturelt rom.

Hay og Penney (2013) bruker «et sosiokulturelt perspektiv» som undertittel på sin bok om vurdering i kroppsøving for å framheve hvordan utøvelse av faget, og vurdering, er med og konstruerer kunnskapsforståelser og «bilder» av elever og hva elever kan. I en nylig utgitt norsk lærebok med både fokus på og tittel *Inkluderende kroppsøving* (Standal & Rugseth, 2015), diskuteres flere slike perspektiver på hvordan ulike praksiser i faget kan være med å forsterke, avkle eller rekonstruere oppfatninger av elever og fagforståelse (se også Svendby, 2016). Det sosiokulturelle kunnskapsperspektivet åpner i avhandlingen for at VfL i kroppsøving kan komme til uttrykk, rekonstrueres og forstås ulikt blant lærere og elever, og at det forskningen, eller læreplan og offentlige dokumenter, sier om VfL igjen kan differere fra dette. Videre inviterer det til et kritisk perspektiv på hvordan implementering og praktisering av VfL kan ha utilsiktede og potensielt negative konsekvenser for enkelte elever. Selv om avhandlingen benytter noen nøkkelstrategier hentet fra tidligere forskning på formativ vurdering for å identifisere VfL praksiser i kroppsøving, understreker den sosiokulturelt informerte virkelighetsforståelse at VfL ikke er noe gitt, ferdig forstått eller et ferdig utarbeidet pedagogisk verktøy. Nøkkelstrategiene kan «pakkes ut» på ulike måter. I tillegg til å kartlegge implementering av VfL fra elev- og lærerperspektiv, søker derfor analysene nye idéer og bidrag til hvordan VfL kan se ut og praktiseres i kroppsøving.

Rorty (1989) starter sin bok *Contingency, irony, and solidarity* med setningen «about 200 years ago, the idea that truth was made rather than found began to take hold of the imagination of Europe» (s. 3). Kunnskapssynet i avhandlingen er i følge dette ingen ny idé, men den ble raffinert og radikalisert i forrige århundre, kanskje spesielt gjennom det som gjerne kalles den lingvistiske vending i filosofien og utviklingen av konstruktivistiske

---

«er historiske og kulturelle konstruksjoner som velges og formes i prosesser der samfunnmessige endringer, politiske ambisjoner og pedagogiske hensyn spiller vesentlige roller» (NOU 2014:7, 2014, s.11).



grunnsyn innenfor sosialvitenskapene<sup>41</sup>. Alvesson og Sköldbberg (2009) skriver at de fleste forskere innenfor sosial- og samfunnsvitenskap i dag vil stille seg bak en forståelse av samfunnet og dets institusjoner som, i en vid forstand, sosialt skapt. De sier at i denne vide betydningen; «most of us are social constructionists» (s. 34), men avgrenser selv begrepet sosial konstruktivisme for en snevrere bruk. Som vist til i kappens tidligere kapitler er VfL tuftet i en slik forståelse av virkeligheten ved å være tett vevd sammen med fremveksten av konstruktivisme og sosiokulturelle læringsteorier i forrige århundre. Men som teoridrøftingene til Black og Wiliam (f. eks. 2009, 2012a; se også kappens kapittel 2) gjør tydelig, kan praktisering av VfL i klasserommet finne forankring i en rekke pedagogiske teorier som i større eller mindre grad også utgjør filosofiske virkelighetsforståelser.

Dette taler for en pragmatisk virkelighetsforståelse der ulike teorier og metodevalg, samt språklige anliggender, betraktes medskapende og kan danne grunnlag for varierte beskrivelser av VfL (og andre forhold) i kroppsøving og skolen generelt. Creswell (2009) uttrykker at, «pragmatism open the door to multiple methods, different worldviews, and different assumptions, as well as different forms for data collection and analysis» (s. 11). Med et pragmatisk utgangspunkt forstår jeg forskningsprosessen som en refleksiv og dialogisk omgang med empirien der interpretasjon eller fortolkning har forrang. Jeg følger her Alvesson og Sköldbberg (2009) i deres utlegning av en refleksiv metodologi, hvor det er et uttalt mål å nedtone betydningen av empiri som beskrivelser av virkelighet og heller betone den som forskerens og teoriens dialogpartner; «as an input to the generation of ideas and the finetuning of reason» (Alvesson, 2011, s. 136).

Deres utlegning er hovedsakelig siktet mot kvalitativ forskning, men de kommenterer selv i en note at «much if not all of what we have to say is relevant to quantitative research. Actually a good deal of the criticism we make touches to even greater extent on the quantitative methods, such as the adoption of a naïve view of language» (Alvesson & Sköldbberg, 2009, s. 14). I utviklingen av refleksive metodestrategier trekker de veksler på såpass ulike forskningstilnæringer som *grounded theory*, postmodernisme, kritisk teori og hermeneutikk. Jeg forstår dem slik at det å utruste seg med forskjellige teorier og tillate en friere bruk av empirisk materiale kan produsere bedre innsikt om fenomener i verden enn ved å søke en så korrekt gjengivelse av innsamlede data som mulig. De skriver at empirisk forskning i refleksivt modus

---

<sup>41</sup> Wittgenstein (2003) og utgivelsen av hans *Philosophische untersuchungen* i 1953 er en sentral kilde i framveksten av det første, mens Berger og Luckmanns (1966) *The social construction of reality* muligens er den mest vanlige opphavsreferansen for sosial-konstruktivistiske metodologier innen sosial- og samfunnsvitenskapene.

... starts from a sceptical approach to what appear at a superficial glance as unproblematic replicas of the way reality functions, while at the same time maintaining the belief that the study of suitable (well thought out) excerpts from this reality can provide an important basis for a generation of knowledge that opens up, rather than closes, and furnishes opportunities for understanding rather than establishes 'truths' (s. 9).

Mye lengre ut i samme bok presiserer Alvesson og Sköldberg (2009) rollen til empiriske data i det reflekseive forskningsperspektivet: Empirisk materiale kan

inspire ideas and theories, endowing them with credibility, clarifying them and, in the case of theories, making them more stringent. Empirical material can not falsify or verify theories, but can generate arguments for or against championing of theoretical ideas and a particular way of understanding the world (s. 303).

Bare unntaksvis vil interpretasjon av et empirisk materiale føre til det de kaller torpedering av en teori; i de aller fleste tilfeller vil empirisk forskning føre til modifiseringer og utvikling av teorien. Herunder kan det grovt skilles mellom reflekseive forskningspraksiser som vektlegger å produsere nye innsikter og de som retter seg mot å løfte fram og unngå det som kan være problematisk eller farlig. Alvesson og Sköldberg (2009) referer til den siste som D-refleksivitet, hvor D står for «deconstruction, defensive, destabilize» (s. 312). Den første kaller de R-refleksivitet med henvisning til «reconstruction, re-presentation» (s. 313). Denne avhandlingens mål om å utvikle og forstå VfL og tilhørende teoretiseringer innenfor en kroppsøvingsramme plasserer den definitivt innenfor R-refleksiviteten. Relatert til kroppsøving og VfL kan Tolgfors og Öhmans (2014) foucaltinnrammede studie av hvordan formativ vurdering og VfL ledsages av underliggende disiplinerte og styrende mekanismer eksemplifisere D-refleksivitet<sup>42</sup>.

Fra å ta utgangspunkt i Giddens (1976) etter hvert velkjente forslag om nødvendigheten av en dobbel hermeneutikk i det mennesker skal fortelle eller rapportere om

---

<sup>42</sup> Inndelingen mellom R- og D-refleksivitet er også grov i den forstand at det å løfte fram det problematiske i D-refleksivt modus ofte betraktes som et steg på veien, og gjøres med mål om, å produsere ny eller forbedret kunnskap. Tolgfors og Öhmans (2014) studie er i så måte ikke entydig, og når jeg lar dem eksemplifisere D-refleksivitet baseres det på at deres studie, sammenlignet med denne avhandlingens studier, har et uttalt kritisk perspektiv på VfL.

egen praksis og forskeren dermed er prisgitt et allerede fortolket materiale, tar Alvesson og Sköldberg (2009) veien om en trippel hermeneutikk før de presenterer en firdrappel hermeneutikk som eksempel på det de velger å kalle «reflexive interpretation» (s. 271). Refleksiv interpretasjon handler om å bevege seg relativt fritt mellom fortolkningsnivåer og arbeide ut fra de spørsmål som er stilt. Dette angir altså den grunnleggende forståelsen av empiriens rolle i avhandlingen. Artikkene kan beskyldes for ikke å gjøre dette tydelig nok, men det er likevel et felles forhold mellom analyse, fortolking og virkelighetsforståelse som er lagt til grunn for valg og slutninger sett under ett. Selv om de kvantitative analysene aksepterer en form for korrespondanseteori, ligger det til den refleksive interpretasjonen at det ikke eksisterer noen direkte kopling mellom «funn» i analysene og det som faktisk utspiller seg i skolene. Det finnes ingen metoder, hverken kvantitative eller kvalitative, som ubesudlet kan hente fram «virkeligheten i skolene». Med dette som utgangspunkt framhever Alvesson og Sköldberg (2009) at refleksiv interpretasjon også betyr at ulike nivå i fortolkningen reflekterer hverandre. Ved gjennomtenkt bevegelse mellom nivåer i fortolkningsnivåene vil det være mulig å gi en tettere beskrivelse av (her:) VfL i skolene. Et eksempel som illustrerer dette er forståelsen av elevenes rapportering av feedback slik det diskuteres i kappens kapittel 5.

Samtidig er det på en måte riktig å si at det ikke er noen skoler i denne avhandlingen. Visst har jeg samlet noen data ved seks skoler. Elever, som vel må regnes som en viktig del av skolene, har fylt ut digitale spørreskjema. Lærerne som underviser i kroppsøving har måttet sitte sammen og prate om egen vurderingspraksis, vel vitende at alt ble tatt opptak av og at jeg kunne finne på å gjengi det de sa i ulike presentasjoner av forskningen. Jeg ser ingen grunn til å betvile at disse lærerne, når de ikke deltok i fokusgruppene med meg, faktisk underviser i kroppsøving og delvis andre fag, er på møter, ringer hjem om elever, etc., men likevel er det jeg måtte ha «funnet» fra disse skolenes praksis å betrakte som noe (i tekst båret fram) annet enn det som «virkelig» utspiller seg i skolene. Avhandlingens «funn» er, med Alvesson og Sköldbergs (2009) ord, «a construction of the empirical conditions, imbued with consistent interpretive work» (s. 284).

Innsamlede data, mine analyser og måten de presenteres på, *skaper* (det jeg igjen med en metafor kan kalle) et bilde av skoler, av lærere og elever. Dette bildet, eller bildene, benyttes til å undersøke VfL i kroppsøving. Det er således strengt tatt VfL i kroppsøving, og ikke skolehverdagen, elevene eller lærerne, som representeres. Men siden analysene og representasjonen har empirisk utspring i seks videregående skoler hvor elever har fylt ut spørreskjema og lærere som underviser i kroppsøving har deltatt i fokusgrupper, kommer jeg,

som jeg har gjort hittil, til å fortsette å omtale utvalget som seks skoler med tilhørende lærere og elever. Poenget er at avhandlingens resultater skal leses i lys av at materialet, de faktisk foreliggende elev- og lærerdata, vil kunne dekonstrueres, analyseres og presenteres på andre måter og ut fra andre teorier og modeller enn de jeg har aktivert, noe som ville ført til andre bilder av skolene og en annen representasjon av VfL. Refleksivitet, skriver Alvesson (2011), «stands for conscious and consistent effort to view the subject matter from different angles, strongly avoiding a priori privileging a favoured one» (s. 106). Derfor har jeg i forskningsprosessen søkt å kartlegge mine fortolknings premisser og fremvise hvordan presenterte kunnskap om VfL i kroppsøving er utviklet og reproduisert gjennom forskningens meningsskapende kontekster.

All datainnsamling ble gjennomført i 2011 og 2012. Metodene som er benyttet i artiklenes studier er beskrevet i respektive artikkel. I tillegg gir presentasjonen av artiklene i kappens kapittel 4 også en kort oversikt over metodetilnærmingen i de ulike artiklene. Jeg ser ingen grunn til å gjenta dette her og vil derfor i det følgende bare utdype noen detaljer og overveielser relatert til planlegging og gjennomføring av fokusgruppeintervju og spørreundersøkelse, samt diskutere studiens forskningsetiske aspekter. Men først skal studiens endelige utvalg beskrives nærmere.

### **Utvalg: Seks skoler, 1486 elever og 23 kroppsøvlingslærere**

Hvem eller hva, og hvor mange? Dette er de primære utvalgsspørsmålene. Det endelige utvalg vil stort sett alltid være bestemt ved en forhandling mellom formål, ressurser, tilgjengelige undersøkelsesenheter og det ideelle utvalg som på empirisk grunnlag ville gi et fyllestgjørende svar på problemstillingen(e).

Til denne avhandlingens studie ble det rekruttert seks videregående skoler. I artiklene brukes begrepet «convenience sampling» om utvalget, da med referanse til Bryman (2012)<sup>43</sup>. Et slikt bekvemmelighetsutvalg, som er tilsvarende begrep på norsk, benytter i prinsippet de enhetene som er lettest å få tak i. Siden rekrutteringen av skoler hadde noen strategiske elementer ved seg er det mer presist, om jeg følger Grønmo (2004), å omtale utvalget som pragmatisk. I den hensikt å sikre bredde i datamaterialet ble det bestemt at studien skulle inkludere både skoler som bare tilbydde henholdsvis yrkesfaglige og studieforbereende utdanningsprogram <sup>44</sup> og skoler som tilbydde begge deler. Ut fra en tanke om at

---

<sup>43</sup> Dette er fjerde utgave. Ved studiens start var det tredje utgave som ble benyttet (Bryman, 2008).

<sup>44</sup> Hvilke navn som har vært brukt for å beskrive de retninger elever kan velge i videregående opplæring har variert. De navn som her er benyttet er de for tiden, anno 2015, gyldige under LK06.

kroppsøvings- og idrettslærerne ved skoler som tilbyr idrettsfag vanligvis vil ha mer organisert faglig samarbeid enn hva som vil være tilfelle for kroppsøvlingslærerne ved skoler uten idrettsfag, ble skoler med idrettsfag ekskludert fra rekrutteringsprosessen. Videre var det et mål å rekruttere skoler av ulik størrelse og med geografisk lokalisering både til Oslo og i et fylke vest i Norge. Det avgjørende utvelgelseskriteriet var likevel skolens interesse i å delta i prosjektet.

Før seks skoler var rekruttert hadde 11 andre takket nei til å delta i studien. Rektor ved en skole takket umiddelbart nei med den klare begrunnelse at han mente elevene allerede var belastet med for mange undersøkelser og kartleggingsprøver fra sentrale utdanningsmyndigheter. Ytterligere åtte av de 11 skolene som takket nei gav en kort grunngiving som hang sammen med at forskningsprosjektet potensielt kunne føre til nye krav til tidsbruk og merarbeid for lærere i en allerede presset skolehverdag. To av de åtte skolene som grunnga sitt avslag refererte også til at de relativt nylig hadde hatt andre forskningsprosjekter som inkluderte kroppsøvlingslærerne. De siste to skolene som takket nei oppgav ingen grunn til hvorfor de valgte ikke å delta i studien.

**Tabell 1. Oversikt over utvalg av lærere og elever ved studiens seks skoler**

	Totalt antall kroppsøvlingslærere	Antall elever som har kroppsøving (N)	Elevsvar inkludert i studien (n)	Svarprosent elever
Skole 1	3	332	283	85,2 %
Skole 2	2	169	103	60,9 %
Skole 3	7	424	306	72,1 %
Skole 4	3	98	87	88,8 %
Skole 5	3	254	173	68,1 %
Skole 6	6	769	534	69,4 %
Total	24	2046	1486	72,6 %

Det er mulig å tenke seg at de skolene som takker ja til deltakelse i en slik studie er mer interessert i enten vurdering eller kroppsøving, eller eventuelt vurdering i kroppsøving, enn hva en gjennomsnittskole ville vært, hvis det hadde vært mulig å snakke om et gjennomsnitt for interesse av slikt mellom skolene. Denne potensielle utvalgsskjevhet utgjør ikke et problem så lenge det ikke er et mål å generalisere resultatene til en bestemt større populasjon, hvilket ikke er et primært anliggende i denne studien. På den annen side, ut over at både

ledelse og kroppsøvingslærere ved skolen altså har sagt ja til å delta i studien, kan jeg tilføye at det ikke er noe ved de data som er innsamlet og analysert som tyder på at disse seks skolene skulle være vesentlig forskjellig fra andre tilsvarende videregående skoler. Tabell 1 gir en oversikt over utvalgets distribusjon på tvers av skolene.

### Beskrivelse av lærerutvalget

Oversikten i tabell 1 viser at det var totalt 24 lærere (9 kvinner og 15 menn) som underviste i kroppsøving ved de seks skolene på tidspunktet for datainnsamling. 23 av disse deltok i fokusgruppeintervjuene, da en lærer (kvinne) var blitt syk på dagen for intervjuet. Det var kvinner og menn i alle fokusgruppene og lærernes alder var fra 24 til 61 år. Fire av lærerne hadde undervist i kroppsøving mindre enn fem år. Alle hadde minst årsstudium i kroppsøving eller tilsvarende som del av sin lærerutdanning eller som etterutdanning. Tre av lærerne hadde hovedfag eller mastergrad i idrett/kroppsøving.

### Beskrivelse av elevutvalget

Tabell 1 forteller at populasjonen av elever som har kroppsøving ved undersøkelsens seks skoler var  $N = 2046$ . 1599 elever fra populasjonen deltok i spørreundersøkelsen (svarprosent: 78,2 %). Etter at materialet var rensket for ikke fullførte spørreskjema<sup>45</sup> og svar gjennomgående uten variasjon i verdier, ble  $n = 1486$  elevsvar inkludert i studien. Kjønnsfordelingen var 841 jenter og 645 gutter. Den litt skjeve fordelingen gjenspeiler en gjennomgående kjønnskjevhet i utdanningsprogrammene som tilbys ved de inkluderte skolene. Siden dette er elever i ordinær videregående opplæring er det bare unntaksvis at elever ikke er mellom 15 og 19 år gamle.

**Tabell 2. Fordeling elevsvar mellom Vg1, Vg2 og Vg3**

trinn	Antall elevsvar	Prosent av alle elevsvar
Vg1	566	38,1 %
Vg2	625	42,0 %
Vg3	295	19,9 %
Alle	1486	100 %

<sup>45</sup> Ens betydende med skjema der en eller flere variabler var «missing». Fordi dette er et digitalt spørreskjema hvor du må svare på spørsmål for å gå videre, betyr «missing» elever som avbrøt undersøkelsen før de fullførte siste spørsmål.

Tabell 2 presenterer elevsvarenes fordeling mellom årstrinnene i videregående opplæring. Det lavere antall deltakere fra Vg3 reflekterer at elever i yrkesfaglig utdanningsprogram avslutter opplæringen i kroppsøving etter Vg2.

Spørreskjemaet (vedlegg 2) inneholder en hel del annen informasjon som kan benyttes til å beskrive elevgruppen. Jeg har valgt å ta med de som direkte kan knyttes til studiene i avhandlingen. Av de elevsvarene som er inkludert i studien rapporterte 524 (35,5 %) at de konkurrerte i idrett, dans eller lignende, mot 956 (64,5 %) som svarte nei til det samme. På spørsmålet «Har du noen funksjonshemming eller kronisk sykdom som gjør det vanskelig å delta i (deler av) kroppsøvingsslaget slik det blir drevet ved din skole?» svarte 256 elever (17,2 %) elever ja, mot 1230 elever (82,8 %) som svarte nei. Med kroniske sykdommer inkludert virker tallet rimelig ut fra en total på 15 % med funksjonshemming i Norges befolkning mellom 16-66 år (Svalund, 2004). Tabell 3 gir en oversikt over det totale elevutvalgets innvandringsbakgrunn. Antall elever med innvandringsbakgrunn er svært forskjellig innbyrdes mellom studiens skoler ut fra geografisk beliggenhet. Ved de tre skolene fra fylket vest i Norge er det få elever som rapporterer innvandringsbakgrunn (25 av totalt 676 elevsvar ved skolene).

**Tabell 3. Fordeling svar på spørsmålet «hvilken oppvekstbakgrunn har du?»**

Svaralternativ	Antall elevsvar	Prosent av alle elevsvar
Jeg er født og oppvokst i Norge	1157	77,9 %
Jeg er adoptert, oppvokst i Norge	40	2,7 %
Jeg har innvandrerbakgrunn (født i annet land)	164	11 %
Jeg har innvandringsbakgrunn (født i Norge, mine foreldre/foresatte er født i annet land)	125	8,4 %
Alle	1486	100 %

Tabell 4 presenterer de karakterene elevene selv oppga å ha mottatt ved siste halvårsvurdering. Sammenligning mellom rapport av karakterer ved de enkelte skolene og skolens faktiske karakteroversikt ved siste halvår viste god overensstemmelse. Jeg noterer også at fordelingen stemmer godt med nasjonal kartlegging (Gravaas, Hægeland, Kirkebøen & Steffensen, 2008), men at gjennomsnittet for 1454 elevsvar på 4,2 er 0,1 til 0,2 poeng

lavere enn der landsgjennomsnittet for sluttvurdering i kroppsøving har ligget relativt stabilt de senere årene<sup>46</sup>.

**Tabell 4. Fordeling av rapporterte karakterer i kroppsøving ved siste halvårsvurdering**

Karakter	1	2	3	4	5	6	iv
Antall elevsvar	12	57	219	510	589	67	32
Prosent av alle	0,8 %	3,8 %	14,7 %	34,5 %	39,6 %	4,5 %	2,1 %

### Fokusgruppene

Fokusgruppeintervju som metode ble utviklet i sammenheng med forskning på propaganda, salg og markedsføring på 1940-tallet<sup>47</sup>. En fokusgruppe kan enkelt beskrives som en gruppe personer som er samlet for å samtale rundt en sak eller tema. Det er vanlig at samtalen ledes av en ansvarlig, en som oftest figurerer under ulike navn som fasilitator, mediator, intervjuer eller ganske enkelt forskeren. Rollen til samtalelederen kan variere, og det er ikke uvanlig med en assistent eller loggfører som hjelper til under gjennomføringen, både med det praktiske og som en ekstra lytter.

De seks fokusgruppene i dette avhandlingsarbeidet ble gjennomført med bare meg som forsker til stede i tillegg til lærerne. De ble gjennomført i et møterom eller klasserom ved den enkelte skole, altså i en fysisk ramme som var kjent for deltakerne. Jeg satte på lydopptaker og innledet samtalen ved å ønske velkommen og repeterte hva vi primært skulle snakke om, men tilføyde at jeg først kunne tenke meg at de sa litt om rammene og status for kroppsøvingfaget på egen skole. Etter denne korte innledningen stilte jeg et åpent spørsmål om de kunne fortelle om vurderingspraksis i kroppsøving ved egen skole. Til hjelp under fokusgruppene var en temaguide laget på bakgrunn av teori om VfL, vurderingsregelverket etter LK06 og indisier fra tidligere forskning i Norge på vurdering i kroppsøving (Vedlegg 4). Temaguiden var spesielt tillempet i den hensikt å sikre at samtalen inkluderte informasjon som kunne fortelle om forståelse og vurderingspraksis relatert til VfL og tilhørende nøkkelstrategier. I intervjusituasjonen prøvde jeg å vente med å introdusere emnene fra temaguiden, i den hensikt å se om lærerne selv ledet oppmerksomheten mot VfL og formative vurderingspraksiser. Det viste seg da også at lærere ved to av skolene selv brukte

<sup>46</sup> SSB statistikk. Nasjonale skolerresultat er hentet fra SSB, Statistisk sentralbyrå, ved søk på nettstedet [www.ssb.no/utdanning](http://www.ssb.no/utdanning).

<sup>47</sup> En oppdatert og kortfattet gjennomgang av fokusgrupper som metode, hva fokusgrupper er og hvordan de har utviklet seg, finnes i Kamberelis og Dimitriadis (2013).



VfL for å omtale deler av hva de prøvde å få til i vurdering i kroppsøving, hvilket igjen ble benyttet som utvalgsriterium i avhandlingens fjerde artikkel (Leirhaug & MacPhail, 2015).

I en intervjusituasjon vil maktrelasjoner i dialogen alltid være et tema, og i fokusgruppeintervju er disse mer komplekse enn i individuelle intervju, fordi de ikke lengre bare angår relasjonen intervjuer og den intervjuede (Hofmeyer & Scott, 2007). Wilkinson (2011) beskriver hvordan samtale «about cases can be interactionally tricky – particular when a presumed ‘expert’ is asking questions, or in settings in which potentially equally knowledgeable others might have different or even conflicting opinions.» (s. 176-177). Begge disse betenkelighetene kan gjelde fokusgruppeintervjuene i denne avhandlingen. Som det sies i artiklene forventet jeg likevel ikke at dette skulle bli et problem, da det å beskrive og diskutere vurderingspraksis må kunne betraktes som en del av lærerens profesjonelle oppgaver. Følgelig skulle ikke temaet være altfor privat (Smithson, 2008) eller problematisk å diskutere foran hverken kollegene eller meg, noe som støttes av min forskermessige evaluering av atmosfæren i fokusgruppene og lesning av det transkriberte materialet. For å dempe det at jeg kunne framstå som en vurderingsekspert, la jeg vekt på å fremme mitt oppriktige mål om å lære mer om faktisk vurdering i kroppsøving og at det var nettopp de, lærerne, som kunne hjelpe meg med dette.

En utilsiktet konsekvens av å velge den eksisterende gruppen lærere som underviser i kroppsøving ved en skole til fokusgruppene, var at det ble kun to og tre lærere i enkelte grupper (se tabell 1). Ut fra diskusjon i metodelitteratur er det et spørsmål om ikke disse mer treffende bør kalles gruppeintervju på grunn av at min stemme som forsker blir mer framtrødende og aktiv i interaksjonene (Bryman, 2012; Parker & Tritter, 2006). Ivanoff og Hultberg (2006) understreker at en viktig forskjell er at interaksjoner mellom deltakere i fokusgruppemetodologi behandles eksplisitt som «research data» (s. 125). Siden dette er tilfellet i analysearbeidet for alle lærerdata, uavhengig av antall deltakere i intervjusituasjonen, virker det, i sammenligning med gruppeintervju som metode, mer presist å opprettholde en vurdering og bruk av alle lærerdata i lys av fokusgruppemetodologi.

Fokusgruppeintervjuene ble spilt inn med en digital lydopptaker (Olympus VN-8600PC) og senere fullstendig transkribert. Transkribering ble gjort av forskeren som også ledet fokusgruppene, altså meg. I praksis innebærer dette arbeidet, slik jeg gjorde det, en begynnende analyse. Jeg satte inn margmerker og skrev ned assosiasjoner og refleksjoner etter hvert som de dukket opp når jeg lyttet til opptak og transkriberte. Teknikken kan beskrives som en systematisk åpenhet for det som sluppmessig dukker opp, og den influerer selvsagt det videre analysearbeidet. Det er også en risiko for at det sluppmessige fra en

tidligere transkriberingsprosess vil farge hva som dukker opp i en senere transkribering. Det ligger også en tilsvarende utfordring i det at det mellom gjennomføring av første og siste fokusgruppeintervju er flere måneder<sup>48</sup>. Begynnende analyse ved transkribering av de første fokusgruppeintervjuene vil kunne påvirke min rolle og ledelse i påfølgende fokusgrupper. For at ikke de første innfall og refleksjoner skal få fortrinnsrett søkte jeg systematisk å overse de sluppmessige notatene i analysens neste trinn. Framgangsmåten var da bruk av en form for tematisk analyse guidet eller informert av teori eller tidligere funn (Braun & Clarke, 2006; Bryman, 2012), der utsagn og tema i transkript fra en fokusgruppe ble kodet og sortert i lys av hele transkriptet og relatert til tilsvarende analyse av de andre fokusgruppetranskriptene. Framgangsmåten er beslektet med forståelsens konstituering i en hermeneutisk sirkel, men grunnes i Alvesson og Sköldbbergs (2009) utlegning av refleksiv metodologi. Teoretisk ramme og videre prosedyrer for analyse er litt forskjellig i de tre studiene hvor lærerdata benyttes (Leirhaug & Annerstedt, 2015; Leirhaug & MacPhail, 2015; Leirhaug, MacPhail & Annerstedt, 2016; se kapittel 4). Redegjørelse for dette og tilhørende overveielser finnes i respektive artikler.

### **Spørreundersøkelsen**

Til spørreundersøkelsen blant elevene ved de seks skolene ble det utviklet et eget spørreskjema. Målsettingen med spørreskjemaet var flersidig, men ut fra avhandlingsstudiets formål ble det utformet for å innhente fyldig informasjon om elevers oppfatning av vurdering i kroppsøving og spisset for å kartlegge erfaring med VfL nøkkelstrategier i vurderingspraksis ved skolen. Samtidig skulle spørreskjemaet kunne gi bakgrunnsinformasjon om elevene og grunnlag for å danne et bilde av deres oppfatning av kroppsøvingfaget mer generelt. Dette siste både for å kunne gi informasjon tilbake til lærerne med tanke på utviklingsarbeid ved egen skole og til bruk som mulige forklaringsvariabler i avhandlingens analyser.

Det foreligger ingen etablerte kartleggingsverktøy for kroppsøving i Norge, og både i artikkel 1 og 2 i avhandlingen (Leirhaug, 2015; Leirhaug & Annerstedt, 2015) vises til at måten å stille spørsmål på var inspirert fra studier av faget *idrott och hälsa* i Sverige (Larsson & Redelius, 2004; Redelius & Hay, 2012). Det ble også kikket på Eriksson m.fl. (2005), og norske spørreskjemaer anvendt i fagdidaktiske kartleggingsundersøkelser blant

---

<sup>48</sup> Det ble lengre tidssprik enn planlagt. Avtalt tidspunkt for to av fokusgruppene måtte flyttes, det ene to ganger. Årsak var justering av fellesplanen ved skolene. Tidsspriket mellom gjennomføring av fokusgrupper er vurdert til ikke å lede til potensielt alvorlige metodologiske svakheter.

kroppsøvingslærere (Arnesen, Nilsen & Leirhaug, 2013; Jacobsen m.fl., 2001, 2006; Vavik m.fl., 2010; Vinje, 2008a). Jeg brukte også min egen erfaringsbakgrunn som kroppsøvingslærer og et spørreskjema laget etter LK06 til en masteroppgave (Wiken, 2011). Det ble lagt vekt på å lage enkle og entydige formuleringer. Spørsmål ble tilpasset begrepsbruken i læreplanen LK06 og det ble utviklet og tilføyd spørsmål som skulle sikre informasjon relatert til VFL nøkkelprinsipper.

Fordi spørreskjemaet skulle gjennomføres digitalt via internett, ble formmessig oppsett og måte å stille spørsmål tillempet stilen i Elevundersøkelsen<sup>49</sup>, et format elevene vil være kjent med fra deltakelse i denne. Utkast til spørreskjema ble diskutert med fagfeller og veileder. En digital pilot ble kjørt med tre klassegrupper i kroppsøving (67 elever) ved en videregående skole som ikke var aktuell for hovedstudien. Piloteringen førte til kun små justeringer av spørreskjemaet, først og fremst forsiktig nedkorting ved å fjerne påstander under hovedspørsmål som elevene opplevde overlappende.

Det endelige spørreskjema besto av 32 hovedspørsmål og hadde 205 variabler (vedlegg 2). Dette er i overkant av hva som er tilrådelig for aldersgruppen, men jeg kunne, etter gjennomføring, konstatere at storparten av elevene virket å ha fullført spørreskjemaet på en god måte. Av totalt 1599 elever som startet utfylling ble 1486 elevsvar funnet gyldig etter at alle med «missing» eller manglende variasjon i verdier var tatt vekk.

Spørreundersøkelsen ble gjennomført i en ordinær time ved elevenes skole. Fortrinnsvis ble andre timer enn kroppsøving benyttet i den hensikt å øke sannsynligheten for svar også fra elever som sjelden deltar eller møter opp i kroppsøving<sup>50</sup>. I de aller fleste tilfeller besøkte jeg, forskeren, elevenes klasserom – eller tok imot elevene på en datalab, hvilket gjaldt for to skoler der elevene ikke hadde bærbar PC –, presenterte undersøkelsen og oppmuntret ærlige svar. På grunn av tidsplanen ved de to største skolene ble hovedandelen av undersøkelsen der administrert av klasselærer som leste opp et informasjonsskriv fra forskeren (samme skriv som var grunnlag for informasjonen der jeg administrerte datainnsamlingen; vedlegg 1).

---

<sup>49</sup> «Elevundersøkelsen» er navnet på Utdanningsdirektoratets årlige kartleggingsundersøkelse blant elever i norsk skole. Den gjennomføres digitalt. Ytterligere informasjon om Elevundersøkelsen finnes via Utdanningsdirektoratets hjemmeside: [www.udir.no](http://www.udir.no).

<sup>50</sup> F. eks. Lyngstad (2012) tar opp denne problematikken, som vel må regnes som en kjent utfordring for mange kroppsøvingslærere. Jeg har selv kommentert dette tidligere (Leirhaug, 2010) med utgangspunkt i en rapport fra *Skolinspektionen* (2010) i Sverige, som gjennomførte uanmeldte besøk i over 300 undervisningsøkter i *Idrott och hälsa*. Tilsynet viste i gjennomsnitt at omtrent hver femte elev ikke deltok i øktene i Sverige, og at det i hver tiende inspiserte økt var 30-50 % av elevene som enten var fraværende eller til stede, men ikke deltakende.

Det ble inngått en databehandleravtale med websurveytjenesten hos Norsk samfunnsvitenskapelig datatjeneste (NSD). Den internettbaserte spørreundersøkelsen ble dermed gjennomført via en nettside hos NSD og innsamlete data ble oppbevart i et elektronisk arkiv hos dem (Vedlegg 6).

Å utvikle et eget spørreskjema medfører også utfordring ut over å stille gode spørsmål og utvikle hensiktsmessige skalaer for svaralternativer. Det gjør at analysene da må tuftes i et ikke-validert verktøy. Hvordan dette er håndtert er beskrevet i artiklene, men det er viktig å understreke at alle resultater fra spørreundersøkelsen må tolkes med forsiktighet. I tillegg kommer de mulige feilkilder som alltid hefter ved selvrapporterte data, selv om et utvalg av denne studiens størrelsesorden vil utligne problemer der det ikke er tale om systematisk under- eller overrapportering for gruppen (Ringdal, 2013). På bakgrunn av de beskrevne reservasjoner på vegne av elevdata ble statistiske analyser utført med eksplorerende målsetninger. De tar ikke sikte på generalisering ut over skolene, tolkes i lys av Alvessons og Sköldbbergs (2009) refleksive metodologi og påberoper seg ikke å ha testet fenomenet eller forskningsspørsmålet ut over hvordan det framtrer i materialet av elevsvar.

Før forskningsetiske sider ved prosjektet kommenteres, vil neste avsnitt utdype noen detaljer rundt utførte statistiske analyser. Grunnlaget for analysevalg tilknyttet enveis ANOVA og post-hoc-tester i artikkel 1 kommenteres kort, men først og fremst utdyper avsnittet forutsetningene for regresjonsmodellen som benyttes i artikkel 2, da disse i artikkelen bare nevnes kort som tilfredsstilt. I samme artikkel brukes også faktoranalyse for å identifisere variabler til regresjonsanalysen, men der er alle vesentlige data med i selve artikkelen. Alle statistiske analyser ble utført med softwarepakken IBM SPSS versjon 21.

### **Bruk av enveis ANOVA og regresjonsanalyse**

I artikkel 1 (Leirhaug & Annerstedt, 2015) benyttes enveis ANOVA med post-hoc-test for å analysere statistiske forskjeller mellom skolene i forhold til variabler selektert på bakgrunn av V<sub>F</sub>L teori. Eta kvadrert ( $\eta^2$ ) ble kalkulert som effektmål og gir et inntrykk av størrelsen på forskjellene mellom skolene. De statistiske data fra analysen er gitt i en appendix til artikkelen. Forutsetninger om normalfordeling i elevdata ved hver skole var møtt, men Levene's test indikerte at variansen i data skolene mellom var signifikant forskjellig. Både Welch's  $F$  og Brown-Forsythe  $F$  kan håndtere dette (Field, 2013). Begge disse ble inspisert ut fra analyse i SPSS, og i artikkelen er det valgt å rapportere Welch's  $F$ . For å finne hvilke skoler som skilte seg ut fra de andre for de ulike variablene ble Games-Howell valgt som

post-hoc-test. Field (2013) anbefaler Games-Howell prosedyren når det er utfordringer med variansen og utvalget i gruppene (antall elevsvar fra de seks skolene) er av ulik størrelse.

I studien i artikkel 2 (Leirhaug, 2015) er hypotesen testet i en regresjonsanalyse. For å kunne si at resultatene fra en regresjonsmodell er til å stole på, må en del forutsetninger være tilfredsstillende. De fleste av disse kan testes statistisk, men man skal samtidig huske på at regresjonsanalyse er ganske robust i forhold til brudd på disse forutsetningene (Field, 2013; Ringdal, 2013). Modellen som ble valgt inkluderte ni uavhengige variabler ut fra en vurdering av teoretisk betydning av variablene i tillegg til at de ytte signifikante bidrag til modellen.  $F$  verdien fra ANOVA tabellen i SPSS indikerte at modellen testet som helhet var signifikant. Korrelasjonskoeffisienten  $R^2$  forteller hvor mye av variansen i den avhengige variabelen karakteriseres som i statistisk forstand kan forklares av modellen, og den benyttes også som et mål på hvor god regresjonsmodellen er. Fordi flere uavhengige variabler vanligvis vil føre til høyere korrelasjonskoeffisient, rapporteres «adjustert  $R^2$ » som tar hensyn til antall uavhengige variabler i modellen. Etter en inndeling av Muijs (2011), som brukes i artikkelen, kan en justert  $R^2 = 0,41$  kalles en modell med moderat tilpasning. Selv om det er noe korrelasjon mellom variabler viste forsøk med inkludering av produktvariabler lite endring av modellen. I lys av at analysene var basert på ikke pre-validerte spørsmålstillinger og variabler ble det derfor besluttet ikke å inkludere dem, og det er de videre forutsetningene for at modellen i artikkelen kunne godtas som presenteres i det følgende.

Forutsetninger om linearitet mellom uavhengige variabler og den avhengige ble inspisert ut fra plot-diagram av  $x$  og  $y$  verdier fra SPSS. Residualenes normalfordeling og homoskedastisitet, forutsetningen om lik varians i residualene, ble likeledes inspisert ut fra normal  $p$ -plot, scatterplot og histogram fra SPSS. Materialet viser noe heteroskedastisitet, men avviket er ikke dramatisk og trenger ikke være problematisk når utvalget er relativt stort (Ringdal, 2013).

Forutsetningen om ukorrelerte residualer betyr i forbindelse med spørreskjema at hva en person svarer ikke skal ha sammenheng med hva en annen svarer. I tversnittdata med tilfeldig utvalg vil dette alltid være tilfellet, men når elever går i samme klasse og har samme lærer er ikke dette gitt. Forutsetningen er enklest å sjekke ut fra en Durbin-Watson test. Det foreligger ikke helt klare regler for når avhengighet kan være et problem, men Durbin-Watson lavere enn 1 og høyere enn 3 regnes som problematiske (Field, 2013). En verdi på 2 viser ingen avhengighet. Durbin-Watson testen for modellen i artikkel 2 viste 1,78. Videre ble fravær av multikollinearitet, i betydningen at ingen av de uavhengige variablene (hverken parvis eller gruppevis) må være perfekt korrelerte, undersøkt med «Variance inflation factor»

(VIF) som er et vanlig mål for dette. Field (2013) skriver at hvis VIF-verdien er under 10 og den gjennomsnittlige VIF-verdien (for alle de uavhengige variablene) ikke er mye over 1 er det sannsynligvis ingen grunn til bekymring, hvilket var tilfellet i denne studien (VIF-verdiene for de ni uavhengige variablene: 1,476; 1,096; 1,094; 1,021; 1,304; 1,615; 1,313; 1,130; 1,056).

### **Etiske overveielser**

Avhandlingen forholder seg til generelle etiske normer og retningslinjer som gjelder all forskning. Mer spesifikt forholder den seg til de forskningsetiske retningslinjer for samfunnsfag, humaniora, juss og teologi, slik de er gitt av *Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora* (NESH, 2006). I og med at det er mennesker og skoler involvert blir dette avhandlingsarbeidet berørt av disse bestemmelsene på flere måter, spesielt noen av punktene som omhandler beskyttelse av personer i forskning (s. 11-21). Det forklares i det følgende hvordan jeg har håndtert de momenter som er mest relevant for denne avhandlingens studier.

Prosjektet ble meldt inn til og godkjent av Norsk samfunnsvitenskapelig datatjeneste (NSD) før oppstart datainnsamling i 2011 (Vedlegg 5). All deltaking i studien var basert på informert samtykke og frivillighet. Lærerne som deltok i fokusgruppeintervjuene signerte et skriv om samtykke til deltakelse, lydopptak og videre forskningsbruk av materialet (Vedlegg 3). I samråd med NSD ble det avklart at det ikke var nødvendig med skriftlig samtykke fra elevene i forhold til spørreundersøkelsen. De fikk derfor, av forskeren (meg) eller en lærer, opplest et informasjonsskriv om undersøkelsen hvor det kom klart fram at det var frivillig å delta og at de kunne trekke seg når som helst uten videre grunngeving (Vedlegg 1). Første (nett)side i spørreundersøkelsen refererte til denne informasjonen og gjorde det klart at du (eleven) ved å «klikke deg videre bekrefter du å ha mottatt denne informasjonen og at du er villig til å delta» (jfr. vedlegg 2).

Personvernet i prosjektet er videre basert på konfidensialitet og anonymisering. Skolene gis enten nummer eller bokstaver i all resultatpresentasjon. Lærere får tilsvarende nummer, med unntak av de tre lærerne i artikkel 4, hvor navnene Geir, Ronald og Leif er pseudonymer. For trygg lagring og behandling av lærerdata ble lydopptak lastet over på passordbeskyttet PC umiddelbart etter gjennomført fokusgruppeintervju, og alle navn og andre referanser som kunne identifisere skole eller lærere ble fjernet/kodet allerede i transkriberingen.

For datainnsamling og trygg lagring av elevdata var det som nevnt ovenfor inngått en databehandleravtale med websurveytjenesten hos NSD. Elevene logget seg inn på spørreundersøkelsen med en kode som kun identifiserte hvilken skole de tilhørte. Dette ut fra at utvelgelsesnivået og analyseinteressen var på skolenivå. Etter fullendt undersøkelse ved en skole kunne en resultatfil lastes fra websurveytjenestens administrasjonsside. Dette ble gjort direkte til en passordbeskyttet PC. Elevdata er i utgangspunktet anonyme ved at de behandles statistisk som gruppe og kun knyttes til skolekode og antall lærere som underviser i kroppsøving ved skolen. Likevel var det lagt inn en buffer om ikke å referere elevdata fra enkeltskoler for variabler med under fem elevsvar. Denne reservasjonen ble det ikke nødvendig å iverksette.

Spørsmål om etikk og troverdighet angår også min tolkning og (re)presentasjon av data, spesielt studiens kvalitative lærerdata. Mine fordommer i forhold til denne undersøkelsen er mange. Bortsett fra en tid i ungdomsskolen, hvor kroppsøving var preget av fotball, sirkeltrening og et sosialt spill hvor jeg ikke passet inn, har jeg som elev hatt et positivt syn på kroppsøving og opplevd mye mestring og gode resultater i faget. Jeg har så langt tilbake som jeg kan huske hatt stor glede av idrett og friluftsliv, og sett på det som en vesentlig ingrediens i det gode liv. Ved å tillegge 11 års erfaring som lærer i kroppsøving, innblandet noen andre fag, hvorav sju av årene i videregående skole, er et nærliggende spørsmål hvordan sikre at fortolkninger og resultatpresentasjon ikke blir for verdiladet og personlig preget. Bryman (2012) skriver at det dominerende perspektivet på dette i dag er at det er uaktuelt for forskere å ha full kontroll på verdiene sine, og at skillelinjene mellom forskningstradisjonene mest går på hvor mye oppmerksomhet dette skal ha i forskningen. Jeg gir det ikke noen spesiell oppmerksomhet, men har, med støtte i den refleksive metodologiens (Alvesson, 2011; Alvesson & Sköldberg, 2009) vektlegging av interpretasjonens forrang og plassering av ansvaret hos forskerens refleksivitet, prøvd å holde kontroll på egne fordommer gjennom å samarbeide om analyse og å prøve ut fortolkninger i diskusjon med andre, da primært artiklenes medforfattere. Tilsvarende har jeg håndtert den konstante utfordring det er å beholde presisjonen i språklige uttrykk ved oversettelse av sitat fra fokusgrupper og item fra spørreskjema. Jeg har selv oversatt fra norsk og prøvd å ivareta uttrykket i de opprinnelige norske sitatene fra fokusgruppene ved blant annet å diskutere oversettelsene med en engelskspråklig fagperson (Ann MacPhail), der jeg fikk anledning til å bruke flere ord for hvordan jeg forsto det norske utsagnet og deretter fikk vurdert hvordan dette var formidlet i oversettelsen.





## Kapittel 4. Presentasjon av artiklene

Kapitlet gir en kort oppsummering av hovedresultatene ved å presentere avhandlingens fire artikler. Artiklene presenteres under sin engelske tittel og kommenteres en etter en i den rekkefølge de foreligger i avhandlingen. Siden de er skrevet på engelsk har jeg valgt å gi rom for litt fylde i de norske resymeene. Presentasjonen gjøres likevel med vekt på å løfte opp forskningsspørsmål og hovedfunn i hver artikkel. Det er primært dette som skal diskuteres opp mot avhandlingens formål og overordnet problemstilling i neste kapittel. I tillegg presenteres kort grunnlagsmateriale og metodetilnærming for hver artikkel.

Selv om det er jeg alene som har skrevet resymeet fra hver artikkel, bruker jeg «vi» der det snakkes ut fra resultater og handlingsvalg i de tre artiklene hvor det har vært medforfattere. Referansebruk er begrenset til et minimum og bare inkludert i den grad de direkte har betydning for presentasjonens innhold. Ut over dette finnes referanser til bakgrunn for forskningsspørsmål og hypoteser og mer detaljert informasjon i selve artiklene.

### **Artikkel 1: Assessing with new eyes? Assessment for learning in Norwegian physical education**

[Av Petter Erik Leirhaug og Claes Annerstedt, publisert online før trykk 14.10.2015 i *Physical Education and Sport Pedagogy*, 1-16.]

Første artikkel har som formål å undersøke om satsingen på VfL og undervisvurdering i LK06 har ført til endringer i vurderingspraksis i kroppsøving. I artikkelen leses LK06 og tilhørende endringer av vurderingsregelverket som et norsk eksempel i en internasjonal bølge av skole- og vurderingsreformer som promoterer eller prøver å implementere formative vurderingspraksiser som VfL. Med Berry (2011) beskrives disse som reformer med en framtreddende VfL agenda, og tolkes i artikkelen som et svar på det Stiggins (2002) overskriftsmessig navnga som en vurderingskrise, etter å ha konstatert et utdanningspolitisk og læreplanmessig fravær av VfL i USA. Situasjonen i Norge etter innføringen av LK06 gir en mulighet for å studere hvilken utbredelse utdanningsmyndighetenes generelle satsing på VfL har fått i praktisk vurderingsarbeid i kroppsøving.

Studien kan beskrives som et *mixed method* design ved at analyser av kvalitative lærerdata og kvantitative elevdata fra seks videregående skoler kombineres og sammenlignes.

Forskningsspørsmålet presiseres til hvordan elevers og læreres oppfatninger<sup>51</sup> av vurderingspraksis reflekterer fire utvalgte<sup>52</sup> nøkkelprinsipper for VfL. Viktigheten av å inkludere data fra lærere og elever ved samme skoler dokumenteres ved at studien avdekker noe inkongruens i deres perspektiver.

Fra spørreskjemaet besvart av 1486 elever hentes ni påstander eller spørsmål, hvorav minst to svarer til hvert av VfL prinsippene. Et eksempel er hvordan respons på påstanden «Jeg får tilbakemeldinger om hvor jeg er i læreprosessen» og spørsmålet «Omtrent hvor mange ganger så langt i dette skoleåret har du fått tilbakemelding fra lærer om hva du må gjøre for å forbedre kompetansen din i faget?» ble plukket ut for å informere VfL prinsippet «læringsfremmende tilbakemeldinger»<sup>53</sup>. Elevdata for de ni variablene analyseres ut fra deskriptiv statistikk og presenteres i frekvenstabeller. Enveis ANOVA benyttes for å avdekke statistiske forskjeller mellom skolene for de ni variablene og VfL prinsippene. Lærerdataba fra seks fokusgrupper er først blitt analysert tematisk med utgangspunkt i de fire VfL prinsippene, for deretter å bli tolket og presentert relatert til resultatene fra spørreskjema og enveis ANOVA.

Resultatene viser at elever ved alle studiens skoler har erfaring med VfL, men samtidig viser de at flertallet av elevene rapporterer liten eller ingen erfaring med tre av de fire nøkkelprinsippene for VfL. Et flertall av elevene rapporterte ikke å være engasjert i vurdering av eget og andres arbeid i kroppsøving og aldri, eller bare en gang, å ha fått tilbakemeldinger fra lærer i forhold til kompetansemålene i kroppsøving. Resultatene fra spørreundersøkelsen blant elevene finner støtte i analysene av lærerdataba ved at kroppsøvingslærernes samtaler og beskrivelser av vurderingspraksis demonstrerte varierende kjennskap til, og dermed bruk av, det som kan gjenkjennes eller rekonstrueres som VfL nøkkelprinsipper. Funnene er sammenfallende med observasjoner av VfL gjort i tidligere forskning på kroppsøving internasjonalt (Georgakis & Wilson, 2012; López-Pastor m.fl., 2013), og viser at det fortsatt er en lang vei å gå før VfL framstår som en integrert del av undervisnings- og vurderingspraksis i kroppsøving. Selv om de fleste lærere i studien, delvis motivert og presset av de nye vurderingsreguleringene og dokumentasjonskrav i LK06, utforsket nye strategier for vurdering i kroppsøving, er ikke disse tydelig forankret i

---

<sup>51</sup> Her var det opprinnelig brukt «students' reports» og «teachers' descriptions and discussions» av vurderingspraksis. Dette ble påpekt i reviewprosessen av artikkelen som uklart på engelsk, og en av reviewerne foreslo å bruke «perspectives» på begge. Forslaget ble etterfulgt i artikkelen, men i avhandlingens norske kappe varierer jeg mellom å bruke ord som erfaringer, eller som her: oppfatninger, og elev- og lærerperspektiv.

<sup>52</sup> Det er fire slike brukt i denne artikkelen. De er basert på Black og Williams (1998a, 1998b) forskning og tidligere brukt av MacPhail og Halbert (2010) i kroppsøvingskontekst. Se også kappens kapittel 2.

<sup>53</sup> I artikkelen heter dette «providing feedback that moves learners forward».

læringsteori eller utprøves i et artikulert perspektiv der å bedre elevenes læring i kroppsøving har første prioritet.

Analyse av elevdata med enveis ANOVA viste statistisk signifikante forskjeller mellom noen skoler for alle respektive variabler (tabell 2 og tabell 3 i artikkelen gir en oversikt over variablene), men post-hoc sammenligninger indikerte ingen av de seks skolene som forskjellige fra de andre når analysen inkluderte alle fire nøkkelprinsippene for VfL samtidig. Ved også å inkludere effektstørrelser (statistiske data er gjengitt i artikkelens appendiks) og visuell analyse av gjennomsnittspot for variablene, vurderte vi tre forskjeller mellom skoler til å være betydningsfulle nok for videre oppfølging. Bare en av disse fant støtte i analyse av lærerdata. Elever ved skole 2 (etter nummer i artikkelen) skilte seg fra de andre ved mer kjennskap til og hyppigere diskusjon av kompetansemålene i læreplanen. Lærerne ved denne skolen uttrykte tydelig fokus på klargjøring av mål i undervisningen, og de beskrev hvordan en del elever var ivrige etter å diskutere målsetting og vurderingskriterier i kroppsøving. Små forskjeller skolene mellom kan tyde på at innføringen av LK06 har medført en noenlunde lik implementering av VfL i kroppsøving på tvers av skoler.

Studien dokumenterer en begrenset implementering og bruk av VfL nøkkelprinsipper i kroppsøving. Vi konkluderer med at utdanningsmyndighetenes framtreddende VfL agenda i LK06 ikke kommer til syne som erfaring i kroppsøving for over halvparten av elevene ved studiens skoler, og at kroppsøvingslærerne sett som en gruppe ikke besitter nødvendig kunnskap om, eller kompetanse i, VfL til å kunne integrere det som en effektiv del av undervisnings- og læringsarbeidet. Det pekes på et generelt behov for å heve det som på engelsk kalles «assessment literacy» blant lærere i kroppsøving.

Samtidig noterer vi at de fleste lærerne virket å ha erkjent et behov for å endre undervisnings- og vurderingspraksis i kroppsøving, og at alle skolene kan sies å stå i en fase hvor vurderingspraksis i kroppsøving er i endring. I diskusjonen løfter vi opp at implementering av VfL er en tidkrevende prosess og at kroppsøvingslærere flest sannsynligvis vil ha behov for faglig støtte og kompetanseutvikling i arbeidet med å iverksette, utforske og vurdere nytten av ulike VfL strategier i egen undervisning. I tillegg pekes det på at forskning vil være nødvendig for å sikre validitet og rettferdighet i de vurderingsformer og praksiser som utvikles. På bakgrunn av de bilder analysene konstruerer av hvert av de fire VfL prinsippene, avsluttes artikkelen med en klar oppfordring om å involvere elevene mer direkte i vurderingsarbeidet i kroppsøving.

## **Artikkel 2: Exploring the relationship between student grades and assessment for learning in Norwegian physical education.**

[Av Petter Erik Leirhaug; publisert online før trykk 23.9.2015 i *European Physical Education Review*, 1-17]

Intensjonen med avhandlingens andre artikkel var å utforske sammenhenger mellom kroppsøvingskarakter og elevenes erfaring<sup>54</sup> med vurdering for læring i kroppsøving. I artikkelen betraktes dette som en konkretisering av forholdet mellom formative og summative vurderingspraksiser. Det norske skolesystemet krever summativ og formativ vurdering av alle elever og i kroppsøving er gjennomføring og ansvar langt på vei delegert til lærernivået og ofte, i praksis, opp til den enkelte lærer.

Både den forskningsbaserte teoretiseringen av VfL og tidligere forskning på implementering av VfL strategier understøtter at elever som erfarer mer VfL vil øke læringsutbytte, hvilket gjør det rimelig å forvente at de også får bedre karakterresultat sammenlignet med elever som erfarer mindre VfL. I forlengelsen av dette ble det formulert en hypotese om positiv sammenheng mellom mer erfaring med VfL i kroppsøving og bedre halvårskarakter<sup>55</sup>.

Artikkelen benytter svar på spørreskjema fra 1454 elever<sup>56</sup> og tester hypotesen i en regresjonsmodell med elevenes rapport av siste halvårskarakter i kroppsøving som avhengig variabel, et tall fra 1 til 6 der 6 representerer beste kompetanse i tråd med karaktersystemet<sup>57</sup>. Fordi spørreskjemaet var utviklet til denne avhandlingens studier og følgelig ikke et pre-validert instrument, ble et utvalg antatt relevante variabler fra datamaterialet utforsket med eksplorativ faktoranalyse. Kaisers kriterium indikerte fire faktorer som alle ble funnet teoretisk meningsfulle og derfor inkludert som uavhengige variabler. En av faktorene lot seg identifisere som studiens fokusvariabel, et konstrukt for erfaring av VfL i kroppsøving bestående av seks items<sup>58</sup>. Med utgangspunkt i tidligere forskning og teori om hva som

---

<sup>54</sup> På engelsk benyttes både «experiences» og «perceptions», og jeg bruker erfaring tilsvarende åpent – ved at det kunne vært benyttet ord som opplevelser og oppfatninger i stedet. I avhandlingen vil det alltid være tale om «rapporterte erfaringer», via et spørreskjema, men siden metode og refleksjoner over begrensninger forutsettes kjent, finner jeg ikke behov for å presisere dette hver gang.

<sup>55</sup> Se om halvårsvurdering og karakter i kappens kapittel 2. Den er en forskriftsregulert del av undervisningsvurderingen.

<sup>56</sup> Av de 1486 spørreskjemasvarene som er inkludert i avhandlingens elevdata hadde 32 markert for ikke å ha fått karakter i faget ved siste halvårsvurdering og ble derfor ekskludert fra analysene i denne artikkelen. Karaktersystemet gir skolene adgang til å gi elever «iv» – for «ikke vurderingsgrunnlag» – når lærer ikke har grunnlag for å vurdere elevens kompetanse i faget.

<sup>57</sup> Forskrift til opplæringsloven (2006), § 3-4.

<sup>58</sup> Tabell 1 i artikkelen viser resultatet av faktoranalysen. De seks item som utgjør konstruktvariabelen, «AfL perception», er i norsk original: 1) Kroppsøvingslæreren har forklart om vurdering for læring, 2) Jeg får tilbakemeldinger om hvor jeg er i læreprosessen, 3) Jeg får være med å vurdere meg selv, 4) Ved vår skole er

påvirker elevenes karakterer ble ni andre uavhengige variabler hentet ut av spørreskjemaet. Eksempler er kjønn, innvandrerbakgrunn og ukentlig fysisk aktivitetsnivå. De av variablene som i separate regresjonsanalyser viste signifikante sammenhenger ( $p < 0,05$ ) med karakter i kroppsøving ble inkludert i multippel regresjonsanalyse med simultant inngang for alle variabler. Statistiske resultater fra ulike modeller ble forhandlet med teori for å velge den modell med mest forklaringspotensiale (Se tabell 2 i artikkelen).

Den valgte modellen inkluderte ni uavhengige variabler og forklarte 41 % av variasjonen i kroppsøvingsskarakterene. Den variabelen som viste det sterkeste bidraget handlet om oppfatning av rettferdighet i lærers karaktervurdering og om elever får den karakter de fortjener. To andre variabler viste også relativt sterke bidrag – elevens rapportering av fysisk aktivitetsnivå i kroppsøvingstimene og opplevelser knyttet til mestring og motivasjon i kroppsøving. Til sammen representerte de tre variablene<sup>59</sup> om lag 31 % av variasjonen i karakterene.

Hovedresultatet er en statistisk signifikant, men svak, negativ sammenheng mellom elevenes oppfatning av vurdering for læring i kroppsøving og karakterene de får i faget. Den negative sammenhengene indikerer en effekt motsatt av hva som var forventet i hypotesen. Elever som rapporterer mer erfaring med VfL vil med andre ord i gjennomsnitt få lavere karakterer sammenlignet med de av elevene som rapporterer mindre erfaring med VfL. Flere mulige forklaringer av dette resultatet diskuteres i artikkelen. Resultatet forteller imidlertid at det ikke er en god sammenheng mellom formativ praksis representert ved VfL og det summative vurderingsresultat i form av karakterer.

For å konkretisere resultatet gjør konklusjonen et poeng ut av at elevenes erfaring med VfL i regresjonsmodellen virker å forklare omtrent like mye av variasjonen i karakterer som det som kan forklares av forskjellen mellom det å være jente kontra gutt i kroppsøving. I Norge, som i mange vestlige land, får gutter i gjennomsnitt litt bedre karakterer i kroppsøving enn jenter. Men i motsetning til kjønn, der det er et skolepolitisk mål å nivellere forskjellen, er det meningen at mer erfaring med VfL skal føre til økt faglig læring og følgelig også bedre karakterresultat. Hvor avhandlingens artikkel 1 viste begrenset implementering av VfL nøkkelpinsipper i kroppsøving generelt, viser denne studien at heller ikke de av elevene som

---

kompetansemålene grunnlaget for vurdering i kroppsøving, 5) Jeg vet hva jeg må gjøre for å få bedre karakter,

6) Elevene er med og vurderer hverandre.

<sup>59</sup> I artikkelen heter disse variablene «teacher's grading», «physical activity level in PE» og «mastery and motivation» hvor den siste er en konstruktvariabel basert på fem item fra spørreskjemaet (se tabell 1 i artikkelen).

– i forhold til andre elever – rapporterer mer erfaring med VfL i kroppsøving får uttelling i form av bedre karakterer.

Det er viktig å være oppmerksom på svakhetene i studien og artikkelen understreker behov for ytterligere forskning før sikrere konklusjoner kan trekkes. Hovedfunnet i studien tyder imidlertid på alvorlige problemer med reliabilitet og validitet ved faglig vurdering i kroppsøving. Slik sett representerer det ikke bare en stor utfordring til utdanningsmyndigheter som promoterer formative vurderingspraksiser, men også til kroppsøvingslærere og vurderingsforskere, om hvordan analysere og forbedre kvaliteten på faglig vurderingsarbeid generelt og karaktervurdering spesielt. I artikkelen sies det at om ikke framtidig pedagogisk praksis i kroppsøving og senere studier kan oppvise bedre sammenheng mellom VfL og den summative karaktervurdering, er det ikke sikkert elever vil bry seg med å la seg engasjere i VfL. Dette vil igjen gjøre genuin implementering av VfL i kroppsøving praktisk talt umulig, da å engasjere elevene i vurderingsarbeidet er et av nøkkelpriinsippene for VfL.

### **Artikkel 3: 'The grade alone provides no learning': Investigating assessment literacy among Norwegian physical education teachers**

[Av Petter Erik Leirhaug, Ann MacPhail og Claes Annerstedt; *Asia-Pacific Journal of Health, Sport and Physical Education* 7(1), 21-36]

Sitatet i artikkelens overskrift er en oversettelse av sitatet i avhandlingens tittel.

Lærerutsagnet gjenspeiler noe av utfordringene i kjølvannet av artikkel 2, samtidig som det viser innsikt og peker i retning av pedagogiske refleksjoner og handlinger som vil være nødvendig for å møte enkelte av disse utfordringene. Dette er karakteristisk for funnene i denne artikkelen som har som formål å utforske «assessment literacy» blant norske kroppsøvingslærere. Presist spørres hvordan kroppsøvingslærernes beskrivelser og drøftinger av vurderingspraksis reflekterer «assessment literacy»?

Begrepet «assessment literacy» kom inn i kroppsøvingdidaktisk teori med Hay og Penneys (2013) bok *Assessment in Physical Education: A sociocultural perspective*, og artikkelen er et konkret bidrag til deres påpekning av «the need for empirical work focusing on teachers' existing assessment literacies» (s. 81). Når denne artikkelen ble skrevet forelå det ikke, så vidt forfatterne bekjent, noen publikasjoner som knyttet empiri til det teoretiske rammeverket for «assessment literacy» i kroppsøving. Senere har dette tilkommet (Dinan Thompson & Penney, 2015), noe som kan tolkes som en støtte til vårt valg om å prøve ut rammeverket på et foreliggende datamateriale fra fokusgrupper med norske

kroppsøvingslærer. Selv om Dinan Thompson og Penneys (2015) artikkel fokuserer på lærere i «primary physical education» – som i Norge vil tilsvare lærere i kroppsøving på trinn 1-7 – noterer vi at deres funn og anbefalinger samsvarer godt med våre analyseresultat fra fokusgrupper med kroppsøvingslærere i videregående skoler.

Metodisk bygger artikkelen på en teoretisk guidet tematisk analyse av lærerdata fra fokusgruppene ved avhandlingens seks skoler. Hay og Penneys (2013) fire elementer<sup>60</sup> for «assessment literacy» i kroppsøving brukes som initierende koding og som et rammeverk i en reflekterende fortolkningsprosess med siktemål om å se hvordan dette kunne informere de empiriske funn.

Analysene viste at alle de 23 lærerne i studien hadde kunnskap om LK06 og at vurdering i kroppsøving var et tilbakevendende diskusjonstema ved skolene. Kjent med nye reguleringer for vurdering etter LK06, og i ulik grad inspirert og utfordret av disse, var lærerne opptatt av å klargjøre vurderingskriteriene og finne måter for å dokumentere tilbakemeldinger til elevene. Samtidig oppviser det analyserte materialet kun et fåtall eksempler der lærere beskriver forsøk på å engasjere elever i vurdering av eget eller medelevers arbeid og læring i kroppsøving, og det avdekkes et gjennomgående fravær av kritisk bevissthet og refleksive betraktninger rundt maktrelasjoner, ulike muligheter og potensielt negative konsekvenser av eksisterende vurderingspraksis. Så selv om kroppsøvingslærerne ved alle seks skoler er i en prosess hvor de prøver ut nye tilnærminger til vurdering i kroppsøving, konkluderer vi med at kroppsøvingslærernes beskrivelser og drøftinger av vurderingspraksis ikke reflekterer alle fire elementene eller sentrale forutsetningene for «assessment literacy».

Analyseprosessen viste det utfordrende å kode og sortere data fra det transkriberte fokusgruppematerialet til bare én av de fire «assessment literacy»-elementene. På bakgrunn av den erfaringen, og det teoretiske poeng at «literacy» er en kontinuerlig forhandlende prosess der en og samme lærer altså kan være vurderingslitterat i en kontekst og illitterat i en annen, påpeker vi i artikkelen et behov for videre diskusjon og konseptualisering av «assessment literacy» i kroppsøving. En mulig vei som pekes ut er å identifisere forutsetninger for «assessment literacy» som et alternativ til Hay og Penneys (2013) elementer.

For å sikre validitet og sammenlignbarhet i et desentralisert skolesystem som det i Norge, er det fundamentalt at lærerne er godt utdannet i vurdering og forstår det å være

---

<sup>60</sup> De fire elementene er «assessment comprehension», «assessment application», «assessment interpretation» og «critical engagement with assessment». En kort presentasjon finnes i artikkelen.

vurderingskompetent – eller vurderingslitterat – som en pågående prosess der forutsetningene for kvalitetsvurdering kontinuerlig forhandles med læreplan og undervisning. Ut fra resultatene i studien framhever vi tre prioriterte fokus med tanke på å heve denne kompetansen blant kroppsøvlingslærerne og sikre forutsetninger for «assessment literacy» i kroppsøvlingslærerutdanningen. For det første er det et akutt behov for praktiske eksempler og pedagogiske modeller for VfL og formativ vurdering i kroppsøving, og videre, hvordan formative vurderinger kan bidra til summative vurderinger i kraft av karaktergivning. For det andre er det nødvendig å skape bevissthet om betydningen av å utvikle elevenes «assessment literacy». Uten en styrking av vurderingskompetansen blant elevene blir det praktisk talt umulig for lærere å gjøre vurdering gjennomsliktig og forståelig. Sist, men ikke minst, indikerer studien at kritisk refleksjon over vurderingspraksis og konsekvenser av vurdering som del av kroppsøvlingslæreres profesjonalitet og faglige læring vil kreve spesiell oppmerksomhet. Artikkelen avrundes med at «assessment literacy» virker å kunne utvikles til et meningsfullt teoretisk rammeverk for dette arbeidet.

**Artikkel 4: ‘It’s the other assessment that is the key’: three Norwegian physical education teachers’ engagement (or not) with assessment for learning**

[Av Petter Erik Leirhaug og Ann MacPhail; *Sport, Education and Society* 20(5), 624-640]

I avhandlingens fjerde artikkel møter leseren (rekonstruerte) vurderingshistorier fra tre kroppsøvlingslærere. Geir, Leif og Ronald (pseudonymer) ble utvalgt fordi de selv, før de var utfordret på det, brukte VfL om noe de gjorde, eller prøvde å gjøre, i eget arbeid med vurdering i kroppsøving. Siden Leif og Ronald er kolleger, arbeider de tre ved to av studiens videregående skoler.

Av de totalt 23 lærerne som deltok i studiens seks fokusgrupper utmerket disse tre seg altså ved å bringe VfL inn i fokusgruppesamtalen før forskeren hadde nevnt eller spurt om det. Artikkelen kontrasterer dette mot en australsk undersøkelse, der Georgakis og Wilson (2012), etter å ha intervjuet 17 kroppsøvlings- og idrettslærere for å lære mer om VfL, konstaterte at lærerne ikke brukte begrepet VfL og at forskernes induktive analyser av intervjuene heller ikke lyktes i å produsere noen kategorier for VfL. I lys av avhandlingens oppbygning presenteres denne fjerde artikkelen som et forsøk på å følge opp deler av de tre første artiklene etterspør. Med resultater som viser at VfL ikke er en svært utbredt praksis i kroppsøving ved de seks skolene, at lærerne viser varierende kjennskap til og forståelse for VfL og at det generelt mangler gode eksempler på iverksetting av VfL i kroppsøving, ble det



interessant å se nærmere på hvordan lærere som (mener de) praktiserer VfL i kroppsøving beskriver og grunngir denne praksis.

Målet er å utforske VfL i kroppsøving gjennom de tre vurderingshistoriene og vi spør hvordan VfL er motivert, forstått og iverksatt av kroppsøvingslærere som prøver å integrere VfL i egen undervisnings- og vurderingspraksis. Metodisk nedtoner vi betydningen av datamaterialet som fokusgruppeintervjuer og følger de individuelle stemmene til Leif, Ronald og Geir gjennom fokusgruppene. Empirien blir brukt som en «dialogpartner» (Alvesson & Sköldbberg, 2009) i en rekonstruksjon av lærernes vurderingshistorier. Analyse og presentasjon av resultat er inspirert av narrativ forskning og de gjenfortalte vurderingshistoriene betraktes som forskningsresultat. Vi mener at historiefortellingen i seg selv bærer et potensial til å få lærere og forskere til å se og reflektere over sine egne (oppfatninger om) undervisnings- og vurderingspraksis i kroppsøving<sup>61</sup>.

Ronald, Leif og Geir kommer med beskrivelser fra kroppsøving som kan knyttes til alle fem VfL nøkkelstrategier<sup>62</sup> og uttrykker klart at de arbeider med å utvikle sin egen og skolens vurderingspraksis. Artikkelen gir ikke et klart svar på om dette engasjementet og tilhørende implementering av VfL kommer av et opplevd behov for å endre kroppsøvingfaget mer generelt, eller om det i hovedsak knytter seg til endringer og nye dokumentasjonskrav i LK06. Lærerne gir uttrykk for at vurdering i kroppsøving er utfordrende og at det på flere måter skiller seg fra andre fag. Leif gjør en konkret sammenligning med matematikk, som han også underviser i. Sikkert er det at de er godt kjent med læreplanen og legger vekt på å klargjøre vurderingskriterier og tydeliggjøre vurderingssituasjoner for elevene. Kompetansemål må brytes ned i mer konkrete og direkte håndterbare læringsmål. Det gjør det enklere både å synliggjøre målene for elevene og å bruke dem som utgangspunkt for meningsfull feedback. Når det kommer til hva de praktisk gjør, beskriver alle tre bruk av en type ferdighetsløyper, eller ferdighetstester, i tilknytning til vurderingsarbeidet, og hvordan de har prøvd ut noen konsepter for egenvurdering. Ved utdypning av eksemplene kommer det fram at de fleste refleksjoner og grunnvinger knytter seg til karaktersetning og tilhørende utfordringer med dokumentasjon. Det kan virke som det

---

<sup>61</sup> Artikkelen hadde opprinnelig en referanse til Barone (1995) på dette punktet. Den ble fjernet i forhandling med ordtelling og en vurdering av at gjenstående referanse Dowling, Fitzgerald og Flintoff (2012) langt på vei ivaretok perspektivet. Til sammenligning gir Barone (1995) likevel en mer direkte og tydelig argumentasjon for hvordan fortellinger kan åpne for nye perspektiver og overtale leseren til å retente egne tatt-for-gittheter.

<sup>62</sup> Denne avhandlingens fjerde artikkel benytter de fem nøkkelstrategiene som også er lagt til grunn i Black og Williams (2009, 2012a) utvikling av teori for formativ vurdering, hvilket gjør at nøkkelstrategiene som gjengis skiller seg fra de fire nøkkelstrategiene som benyttes i avhandlingens to første artikler. Utviklingen av VfL er behandlet i kappens kapittel 2.

summative får dominere over de formative intensjonene, og at de blir mer opptatt av vurdering og å klargjøre kriterier enn av å planlegge og tilrettelegge for elevers læring. Deler av det de tre lærerne beskriver som VfL ser mer ut som vurdering for vurderingens skyld.

Hva og hvordan effektiv og velintegrert VfL i kroppsøving kan se ut, vet vi med andre ord fortsatt lite om. Vurderingshistoriene til Geir, Ronald og Leif viser behov for at lærernes utforskning av VfL forankres i læringsteori. I artikkelen vises det til at dette kan være enklere sagt enn gjort. Mens læring og læringsmål koblet til kroppsøving og bevegelsesaktivitet ofte framtrer praktisk og persontilknyttet, er eksisterende teorier om læring i hovedsak kognitivt orientert. Videre har tidligere forskning vist hvordan utdannings- og læringsdiskurser i kroppsøving utfordres og infiltreres av andre sterke samfunnsdiskurser, da spesielt idrettsdiskursen og konkurranseidrettens logikk, samt hvordan det å være i fysisk aktivitet framtrer med kraftige bånd til spørsmål rundt livsstil og helse. På bakgrunn av dette og påvisning av avgrensede forståelse og iverksetting av VfL hos tre lærere som refererer egen praksis som VfL, avrundes artikkelen med behov for mer utprøving og forskning for å kartlegge og forstå mulige konsekvenser av VfL i kroppsøving.

## Kapittel 5. Sammenfatning og veier videre

Den overordnede problemstilling som har ledet arbeidet med avhandlingen er hvordan VfL – vurdering for læring – forstås og implementeres i kroppsøving? VfL er operasjonalisert ut fra nøkkelstrategier primært basert på forskningen og teoriutviklingen til Black og Wiliam (1998a, 1998b, 2009, 2012a) og forklart i kappens kapittel 2. Problemstillingen knytter seg til et økt fokus på vurdering og promotering av VfL som fulgte skolereformen Kunnskapsløftet i 2006. Gjennom fire studier, publisert i ulike tidsskrifter og presentert i kappens kapittel 4, er deler av den overordnede problemstillingen undersøkt både fra elev- og lærerperspektiv ved seks norske videregående skoler. Dette kapitlet har som mål å sammenfatte resultatene og gjøre det eksplisitt hvordan avhandlingens fire studier henger sammen, samt peke på avhandlingens bidrag og videre forsknings- og utviklingsbehov. Selv om det underveis gjøres sammenligninger med og trekkes veksler på internasjonal teori og forskning, tar kapitlet utgangspunkt i at undersøkelsens empiriske materiale er fra norsk videregående skole og vil orienteres mot denne kontekst. Som del av dette vil jeg avslutningsvis diskutere resultatenes betydning i lys av utviklingstrekk i norsk skole og faget kroppsøving.

Avhandlingens artikler omtales i det følgende med nummer ut fra den rekkefølge de foreligger i avhandlingen, i hvilken de også ble presentert i forrige kapittel, altså med benevnelse artikkel 1, artikkel 2, artikkel 3 og artikkel 4<sup>63</sup>. Hensikten er å gjøre det enklere å skille avhandlingens artikler fra andre referanser i teksten.

### Avhandlingens bidrag og artiklenes sammenheng

Først av alt bidrar denne avhandlingen til å fylle et generelt kunnskapshull hva gjelder forskningsbasert kunnskap i og om kroppsøvingfaget i Norge (Jonkås, 2010; Lyngstad m.fl., 2011). Den gjør dette med fokus på forståelse av vurdering og vurdering for læring i kroppsøving i videregående skole. Ved å undersøke dette i lys av LK06 og tilhørende vurderingsregelverk forteller resultatene samtidig noe om innføringen av Kunnskapsløftet og implementering mer generelt. Det var også et mål å bidra inn i internasjonal forskningsdiskurs om vurdering i kroppsøving. Avhandlingsartiklene representerer de første forskningsresultater som er publisert på engelsk i reviewbaserte tidsskrifter fra norsk utdanningskontekst hvor hovedfokus er vurdering i kroppsøving. I så måte kan de vel allerede

---

<sup>63</sup> Artiklenes referanser er Leirhaug og Annerstedt (2015) for artikkel 1, Leirhaug (2015) for artikkel 2, Leirhaug, MacPhail og Annerstedt (2016) for artikkel 3 og Leirhaug og MacPhail (2015) for artikkel 4.

sies å oppfylle uttalte mål, men jeg mener artiklene i tillegg har bidratt til å utpeke viktige innsatsområder og behov for lærerkompetanse og fagdidaktisk utvikling relatert til VfL som del i kroppsøvingens potensielle framtider.

Før jeg vender tilbake til aktualiserte innsatsområder og utfordringer vil jeg starte med en kort gjengivelse av sammenhengen mellom artikkelstudiene: Med utgangspunkt i elev- og lærerdata fra de seks videregående skolene undersøker artikkel 1 implementering av VfL i kroppsøving. Resultatet viser at VfL-agendaen som fulgte LK06 har hatt en effekt i kroppsøving, men indikerer samtidig at hva denne effekten innebærer kan variere og at den faktiske integreringen av VfL nøkkelstrategier i kroppsøvingen sett under ett er avgrenset. Selv om alle kroppsøvingslærerne ved skolene er lærer- og fagutdannet, så er deres beskrevne bruk av, så vel som kjennskap til, VfL nøkkelstrategier varierende. Studien avdekker videre at over halvparten av elevene rapporterer liten eller ingen erfaring med en eller flere av VfL strategiene, hvilket gjør det rimelig å hevde at utdanningsmyndighetenes promotering av læringsfremmende vurderingsstrategier ikke når ut til en relativt stor gruppe elever. På den annen side er det jo elever som rapporterer mer erfaring i forhold til de undersøkte VfL strategiene. Ut fra teori og tidligere forskning<sup>64</sup> kan det forventes at denne mererfaring av VfL støtter det faglige læringsutbyttet for de elevene dette gjelder. Artikkel 2 tester først et teoretisk konstrukt for VfL statistisk og finner det meningsfullt å omtale og studere elever som rapporterer mer erfaring med VfL på tvers av nøkkelstrategier. Der artikkel 1 viser at implementering av VfL ikke er en realitet sett fra mange elevers perspektiv utforsker artikkel 2 videre om det er en sammenheng mellom faglig resultat og de elever som rapporterer mer erfaring med VfL i kroppsøving, det siste uttrykt ved konstrukt for VfL. Artikkel 2 bruker elevenes rapporterte kroppsøvingsskarakter som et uttrykk for faglig resultat og studien viser som kjent at VfL ikke gir utbytte i form av bedre karakterer for de elevene som rapporterer mer erfaring med VfL.

Resultatene i artikkel 1 og artikkel 2 demonstrerer at ikke alt er som det burde være i forhold til VfL i kroppsøving ved studiens skoler. Dette er tolket som indikasjoner på relativt store utfordringer i forhold til validitet og rettferdighet, for øvrig en indikasjon som har god støtte i annen forskning som belyser vurdering i kroppsøving i Norge (Arnesen, Nilsen & Leirhaug, 2013; Dalen & Aune, 2013; Engvik, 2013; Prøitz & Borgen, 2010; Sandvik m.fl., 2012; Vinje, 2008a). Fra svensk forskning på vurdering i kroppsøving minner Svennberg,

---

<sup>64</sup> Mer om forventningens teoretiske bakgrunn finnes i artikkel 2 og kappens kapittel 1 og 2. Forventningen, og hypotesen som konkret undersøkes i artikkel 2, er forankret i de samme kunnskaps- og læringsperspektiver som ligger til grunn for satsingen på VfL i LK06.

Meckbach og Redelius (2014) om at det er en forutsetning at lærerne har god vurderingskompetanse og forståelse for at et kriteriereferensielt vurderingssystem skal kunne fungere rettferdig i et desentralisert skolesystem som det norske og svenske. Studien i artikkel 3 retter følgelig fokus mot kroppsøvlingslærerne ved skolene og bruker en teori om «assessment literacy» i kroppsøving (Hay & Penney, 2013) som ramme for å studere deres beskrivelser og forståelse av vurdering. Med «assessment literacy» som teoretisk ramme identifiseres utfordringer relatert til lærernes vurderingskompetanse, vurderingspraksis og kritisk refleksjon knyttet til vurdering. En del av bildet som framtrer gjennom analysene i artikkel 3 utdyper funn fra artikkel 1 og demonstrerer til dels stort sprik i kjennskap til og forståelse av VfL blant kroppsøvlingslærerne. For å få mer kunnskap om VfL i kroppsøving og kroppsøvlingslærernes forståelse av VfL blir det interessant å undersøke hvordan lærere som uttrykker kjennskap til VfL, og som har gjort seg erfaringer med det i egen kroppsøvlingsundervisning, forstår og beskriver dette. Dette er hva studien i artikkel 4 utforsker ved å rekonstruere og diskutere vurderingshistoriene til tre lærere som selv omtalte noe av det de gjorde, eller prøvde å gjøre, i kroppsøving som VfL.

### **Forskjeller mellom og i skolene**

Avhandlingens forskningstilnærming var designet for å gjøre det mulig å påvise forskjeller relatert til forståelse og implementering av VfL i kroppsøving skolene imellom, og mellom elev- og lærerperspektivet. I forhold til det første viser artikkel 1 at det er ingen av skolene som skiller seg ut når materialet analyseres på tvers av VfL nøkkelstrategier. Når det i tillegg er bare i forhold til framoverrettet feedback som nøkkelstrategi det avdekkes tydelig sprik mellom lærer- og elevperspektivet, og dette også framtrer som gjennomgående for de seks skolene, kan resultatene tyde på at innføringen av LK06 med tilhørende VfL-agenda har hatt en omtrent lik effekt på kroppsøving ved den enkelte skole<sup>65</sup>.

Studien demonstrerer samtidig, jeg vil si til dels sterk, variasjon lærerne mellom når det kommer til hvor inngående kjennskap de viser til den nye vurderingstenkningen, de formative perspektivene som underveisvurderingen bygger på og VfL i særdeleshet. Det er klare forskjeller i hvilken grad studiens kroppsøvlingslærere uttrykker aktiv bruk av læreplanen og hvordan de har oppfattet, vektlegger og operasjonaliserer kompetansemål og vurderingsreguleringer. Artikkel 3 viser hvordan dette også varierer mellom lærere ved

---

<sup>65</sup> I streng metodisk forstand gir forskningsdesignet ikke grunnlag for å trekke konklusjoner om hvor mye effekt LK06 har hatt, da eventuelle forskjeller skolene mellom før innføringen av LK06 ikke er kjent. Men måten lærerne omtaler hvordan de har endret eller holder på å endre vurderingspraksis ut fra fortolkning av LK06 og det nye vurderingsregelverket, gjør det likevel rimelig å spekulere under formuleringen «kan tyde på».

samme skole. Sammenlignet med den generelle følgeforskningen for LK06 (Aasen m.fl., 2012; Møller, Prøitz & Aasen, 2009), kan det virke som det implementeringsarbeidet som der beskrives til i stor grad å ha foregått i lokale kulturer, ved skoler eller i små regioner, er ytterligere differensiert, tidvis helt ned til lærernivå, hva gjelder faget kroppsøving. Dette viser at den enkelte lærer har stor autonomi i sin profesjonsutøvelse, men vitner samtidig om et ekstremt behov for mer samarbeid mellom og oppfølging av kroppsøvingslærerne. Selv om kroppsøvingslærerne har organisert samarbeid ved et par av skolene, er det langt igjen før det er snakk om at de jobber med læreplan og vurderingspraksis i et tolkningsfelleskap av typen som Engvik (2013) mener må til for å etablere en god vurderingskultur. Flere av lærerne ved avhandlingens skoler, i likhet med informantene til Prøitz (2013), peker på tid som knapphetsfaktor og at de derfor prioriterer bort et samarbeid rundt vurdering som de samtidig vet er viktig. Dette kan tolkes som en påminner om at implementering av VFL i kroppsøving ikke bare handler om kroppsøvingfaget internt, men også om prioriteringer og ressurser både lokalt og nasjonalt, om videreutdanning, arbeidstid, budsjett, reiseogtdgjørelse og andre forhold som kontinuerlig innrammer skolehverdagen.

### **Noen hovedtrekk, utfordringer og begrensninger**

Både i artikkel 1 og artikkel 3 gjøres et poeng ut av at alle lærere i studien kjenner LK06 og de mest sentrale endringene ved vurderingsregelverket. Ved alle seks skolene arbeides det med å endre vurderingspraksis i kroppsøving, noe flere lærere grunngir med forankring i forskriften for vurdering. Innføringen av LK06 virker å ha ført til en generell bevisstgjøring omkring læreplan og opplevd behov for å endre (noe i) arbeidet med faglig vurdering. Det er ikke slik som funn i tidligere forskning på implementering i England, hvor læreplanendringer fram mot 2000 hverken førte til skifte i kroppsøvingslæreres filosofi eller praksis (Curtner-Smith, 1999; Green, 2008). I en nyere studie av karaktersetning blant kroppsøvingslærere i Sverige, hvor vurderingsutfordringer i kroppsøving generelt har vært bedre kartlagt enn i Norge, påviser og kommenterer Seger (2014) en utvikling tilsvarende den i Norge. I motsetning til andre svenske studier hvor lærerne har vist til en form for intuisjon eller magefølelse ved karaktervurdering (Annerstedt & Larsson, 2010; Svennberg, Meckbach & Redelius, 2014), viser lærerne i hennes studie til sin profesjonalitet og den nye læreplanen som ble innført i Sverige i 2011. At arbeid med endring av vurderingspraksis forankres i LK06 og nye formuleringer om undervisningsvurdering betyr likevel ikke at utprøving av nye former for vurdering er i tråd med reformens VFL agenda.

Det er flere eksempler i artiklene på hvordan kroppsøvlingslærernes tenkning og uttrykk om vurdering reflekterer et tradisjonelt perspektiv der vurdering primært handler om det summative og brukes nærmest synonymt med karaktersetting. Selv om Sandvik m.fl. (2012; se også Sandvik & Buland, 2014) fant at formative vurderingspraksiser var mindre integrert i kroppsøving sammenlignet med engelsk, norsk og matematikk, er kroppsøvlingslærerne ikke alene om utfordringene relatert til summative og formative vurderingspraksiser og krav. Det er trukket fram som en utfordring for LK06 generelt (Aasen m.fl., 2012) og noe Tveit (2014) beskriver som uavklarte spenninger i det norske utdanningssystemet som igjen plager utvikling av bedre vurderingspraksiser. I England observerte Black m.fl. (2003) noe tilsvarende i en studie som direkte henter erfaring med implementering av VfL. De skriver at det var «evident that the context of national or local requirements for certification and accountability exerted a powerful – usually harmful – influence on assessment practice» (s. 13).

Utdanningssystemets spenninger mellom det formative og summative kan virke å utfordre kroppsøving på linje med andre fag selv om det hverken er nasjonale prøver eller avsluttende eksamen i faget. Derimot er det, som Prøitz (2013; se også Prøitz & Borgen, 2010) påviser, sannsynligvis viktig å akte fagspesifikke utfordringer når vurdering tolkes gjennom norske kroppsøvlingslæreres øyne, for øvrig støttet av en forskningsmessig forventning om at fagtradisjon innvirker på læreres vurdering (Resch, 2009). En av de fagspesifikke tradisjoner som plager vurderingspraksis i kroppsøving er den fremtredende posisjonen ulike tester av fysisk form, prestasjon og ferdigheter har hatt<sup>66</sup>. En mindre undersøkelse gjennomført av Elnan og Sando (2014) er egnet til å illustrere. De observerte tre strategisk valgte kroppsøvlingsøkter hvor elevenes kompetanse i kroppsøving ble vurdert ved bruk av tester med forhåndsgitte kriterier for måloppnåelse. Ut fra observasjonene og kvalitative intervju med fire elever i etterkant, fant de at mange av elevene ikke opplevde bevegelsesglede:

---

<sup>66</sup> Ved endringene av læreplan i kroppsøving i 2012 kommenterte Utdanningsdirektoratet (2012) bruk av tester slik: «Det har vært en tradisjon i kroppsøvlingsfaget å benytte ulike typer fysiske og tekniske tester for blant annet å måle styrke, spenst og utholdenhet. Hverken gjeldende eller revidert læreplan i kroppsøving, eller forskrift til opplæringsloven gir metodiske anbefalinger om bruk av fysiske tester eller testing. Bruk av tester kan være problematisk og være i strid med prinsippene for vurdering. Er tester en egnet metode til å vurdere elevenes kompetanse, slik kompetanse er beskrevet i læreplanen for faget? Kompetansemålene er ikke formulert på en måte som angir at eleven skal vise styrke, løpe fort eller hoppe høyt.» (s. 7). López-Pastor m.fl. (2013) sin review av forskning på alternative vurderingspraksiser i kroppsøving viser at dette med «physical fitness tests» som grunnlag for vurdering også har vært en utfordring internasjonalt i kroppsøving.

Elevenes tydelige signal om at de mistrives med aktiviteten ble ignorert av lærerne. Spesielt betenkelig er det at elevene som sannsynligvis drev minst med fysisk aktivitet på fritiden, var de som responderte mest negativt på undervisningen som ble gjennomført. Det kan se ut som fagets formål i praksis må gi tapt mot andre krav og forventninger som stilles. I vår undersøkelse så det ut til at vurderingsproblematikken og dokumentasjonskravet som lærerne følte på, overskygget formålet med faget (Elnan & Sando, 2014, s. 79).

Det nevnes i artikkel 3 at fem av skolene praktiserer en eller annen form for tester, mens de ved den siste skolen har sluttet å bruke tester. En lærer ved den siste skolen understreker at det «står jo ingen plass at vi skal teste elevene». Ved de andre skolene varierer både vektlegging, form og bruk av testene. Eksempler på hvordan tester er tilrettelagt og forsøkt knyttet til VfL går fram i artikkel 4, mens det i materialet også er eksempler på enkeltlærere som knytter test til karaktersetning og uttrykker praktisk talt et rent summativt perspektiv:

Hvis du seier at det er ein elev eg vurderar til fire, og kanskje, hvis han hadde vist at han kunne sprunge til ein sterk femmer og kan ta armhevingar til ein sterk femmer, men hvis han viser ikkje dette så kan ikkje eg ta det med i vurderinga. Eg kan ikkje anta at han har bra nok kondis. Eg kan sjå at han har bra kondis, men kor bra han er, til fire eller fem, det kan ikkje eg anta eller vite noko om før han eventuelt har vist det på ein Cooper.

Sitatet representerer likevel et unntak i avhandlingens materiale ved at det ikke knytter seg til kompetansemålene i læreplanen og derfor blir et klassisk validitetsproblem. Hovedinntrykket, i likhet med hva Bach (2015) antyder i en relativt fersk masteroppgave, er en bevegelse i retning av at kroppsøvlingslærerne ved de seks skolene har utvidet tilfanget av vurderingsformer etter innføringen av LK06. Men samtidig med at flere lærere prøver ut nye og mer varierte vurderingsmetoder, er utprøvingene ikke systematiske og virker for enkelte lærere mer å være motivert ut fra opplevde krav til dokumentasjon og det å kunne grunngi karakterer, enn at de knytter det til elevenes læring og forståelse av faget.

Variasjonen i oppfatning og forståelse av vurderingsregelverkets krav og VfL blant lærerne, og det at beskrevet praksis med VfL i artikkel 4 fremstår noe avgrenset selv blant lærere som uttrykker god kjennskap til VfL, kan være med å forklare de varierende – men gjennomgående lave – elevrapporteringer av VfL erfaringer i artikkel 1 og



hypoteseforkastelsen i artikkel 2. Det er også andre mulige forklaringer til resultat i artikkel 2 som ikke fikk plass i artikkelen. For eksempel antyder studien til Prøitz og Borgen (2010) at lærerne bruker ulike kriterier som grunnlag for ulike karakterer, og noe tilsvarende kan være tilfelle relatert til VfL nøkkelstrategier. Det kan være slik at elever med lavere karakterer i kroppsøving faktisk får mer VfL, og at de som allerede gjør det bra karaktermessig opplever mindre tilbakemelding og at deres interesser er godt ivaretatt i faget. Dette gir i så fall grunn til å problematisere rådende forståelse og bruk av VfL nøkkelstrategier for høytpresterende, og kanskje godt selvregulerte, elever.

I en Foucault-inspirert analyse av VfL som styringsredskap i kroppsøving i Sverige understreker Tolgfors (2014) at den disiplinerende og styrende effekten kan bli så sterk at det gir grunnlag for å trekke i tvil i hvilken grad dette vil bidra til bevegelsesglede og aktiv livsstil på lang sikt, et perspektiv som er nedfelt i formålet for faget kroppsøving i LK06 samtidig som det er forskningsmessig svakt dokumentert at faget kroppsøving gjør jobben og har denne betydning (Bailey m.fl., 2009; Green, 2014; se også artikkel 4). Tolgfors (2014) sin relevans til resultatet i artikkel 2 er hvordan noen elever, i møtet med VfL som styringsredskap, kan velge «att göra motstånd mot denna styrning och risikerar då att få betala priset genom ett lägre betyg» (s. 152).

Slik jeg ser det, er både Tolgfors (2014) sitt poeng, variert praksis kroppsøvlingslærerne mellom og dette med ulik vurderingspraksis langs karakterskalaen, plausible delforklaring til det teori- og hypotesestridige resultatet i artikkel 2. Dette gir grunnlag for å sette fokus på det metodiske ved rene kvantitative studier som finner sammenhenger. Det er derfor ikke grunnlag for å trekke slutninger fra analysene i artikkel 2 til at kroppsøvlingslærernes vurderingsarbeid ikke holder validitet, spesielt hva gjelder innholdsvaliditet. Det er derimot klart, slik det konkluderes i artikkelen, at det må være en sammenheng mellom VfL og bedre karakterresultat for at elevene skal gidde å engasjere seg. Videre må det utvikles gode eksempler på didaktisk praksis som refleksjonsgrunnlag og tydeliggjøring for lærere hvordan formativ vurdering bidrar til (å lette opplevde utfordringer med) karakter- og sluttvurdering. Hvis ikke, er det sannsynlig at lærerne vil oppleve dette som en tidstyv heller enn et «smartere» pedagogisk virkemiddel, og da blir det vanskelig å argumentere for implementering av VfL og økt fokus på formativ vurdering.

Forskningsdesignet gir ikke muligheter til å studere VfL for den individuelle lærer inn mot sine elevgrupper. Dette var basert på et grunnlagt valg som ble gjort relatert til å studere VfL på skolenivå, men representerer samtidig en svakhet ved studien som er blitt ytterligere aktualisert gjennom påvisningen at det kan være like store forskjeller i kjennskap til og

forståelse av VfL mellom kroppsøvingslærere ved samme skole som mellom lærere fra ulike skoler. Videre ville det vært spennende og forståelsesmessig utfyllende, og er ikke minst på grunn av antydde metodeutfordringer et fremtidig behov for, å undersøke elevenes perspektiv mer inngående enn hva spørreskjematilnærmingen gir anledning til. Aktuelle tilnærminger kan være å gjennomføre fokusgrupper også med elever, eller gjennom å se på praksis i kroppsøving mer direkte for eksempel ved observasjon eller bruk av filmopptak, det siste en metode som er benyttet i Sverige og beskrevet av Quennerstedt m.fl. (2014).

En mer inngående studie av VfL i kroppsøving må også inkludere undersøkning av elevenes læringsutbytte. Både teoribygningen som fundamentierer VfL og utdanningsmyndighetenes agenda for å implementere VfL bygger på hypotesen om at VfL fører til mer og bedre læring for elever. Denne avhandlingen har kun studert elevers rapporterte erfaringer med vurdering og VfL strategier i kroppsøving, ikke elevenes læringsutbytte. Dette gir anledning til å trekke konklusjoner om implementering av VfL i tråd med avhandlingens problemstilling, men det gir ikke grunnlag til å felle beslutninger vedrørende validiteten av de samme (eventuelt implementerte) VfL praksiser<sup>67</sup>. Videre må foreliggende resultater leses i lys av at et utvalg på seks videregående skoler ikke gir forskningsmessig grunnlag for å trekke konklusjoner om andre skoler enn de seks som er undersøkt. Denne enkle metodiske innsikt er likevel ikke helt rett fram. Alvesson og Sköldbørgs (2009) refleksive metodologi, godt posisjonert i avhandlingens pragmatiske og sosiokulturelt informerte virkelighetsforståelse, betrakter resultatene som resultat av analytisk og fortolkende dialog mellom teori, forsker(e) og empiri. Dermed vil en forskningsprosess preget av refleksjon over tid og gjentatte analyser av et empirisk materiale, samlet blant elever og lærere ved seks skoler, på en måte konstruere (bilder av VfL ved) seks skoler. Ved å være rigid, og ikke følge det pragmatiske ved virkelighetsforståelsen, ville jeg kunne argumentere for at avhandlingens seks skoler bare finnes mellom permene i avhandlingen. Jeg kommer ikke til å forfølge denne vesle vitenskapsteoretiske avstikker videre, men den ble gjort for å minne om at det er representasjonen av VfL i kroppsøving, og ikke skolene, som er målet i avhandlingen. Gitt dette som utgangspunkt, har studiene i avhandlingens fire artikler demonstrert at det er sider ved (implementering av) VfL i kroppsøving som har store forbedringspotensialer. Av de VfL nøkkelstrategiene som er undersøkt gjennom avhandlingen avdekkes særlige utfordringer knyttet til (framoverrettet) feedback og involvering av elevene i vurderingsarbeidet, samt utvikling av kroppsøvingslærernes kritiske

---

<sup>67</sup> Se kappens kapittel 2 for en redegjørelse av validitetsbegrepet knyttet til formative vurderingspraksiser og VfL.

refleksjon over eksisterende og planlagt vurderingsarbeid, fortrinnsvis som ledd i kontinuerlig profesjonell utvikling der flere lærere deler og moderer sine erfaringer i et fagfellesskap. Det er disse innsatsområdene jeg vil diskutere i neste to avsnitt før jeg til slutt ser nærmere på framtidsutsikter for VfL i kroppsøving i norsk skole.

### **Feedback i kroppsøving**

Framoverrettet feedback er sentralt for å lykkes med VfL. I oppsummeringen av et arbeid med implementering av VfL nøkkelstrategier blant lærere i England noterte Black m.fl. (2003) at «the quality of feedback provided is a key feature in any procedure for formative assessment.» (s.15). Black og Wiliam (2012a) skriver senere at feedback i klasserommet handler om lærere, elever og fag. For å understreke det kontinuerlige og gjensidige ved feedback som en formativ vurderingsstrategi taler de like ofte om klasseromsdialog som feedback, og det er vesentlig i et fagdidaktisk perspektiv at dette også innebærer en dialog, eller interaksjon, mellom elev og fag. Det kan være et mål for lærer å få eleven til å interagere mer selvstendig med fagkunnskap og læringsmål, noe som vil kunne gjøre elevens videre læringsprosess mindre avhengig av tilbakemelding fra lærer og andre.

Black og Wiliam (2012a) understreker med dette feedbackperspektivets fagspesifitet på en måte som kan gjenkjennes i flere av kroppsøvingslærernes beskrivelser av hvordan de arbeider. I artikkel 3 siteres en av lærerne som sier:

...en sånn litt spesiell ting med kroppsøving, er jo at det er aktivitet, det er ferdighetsprøving og –læring. Mye av det trenger du ikke sette så mye ord på, en kan tilrettelegge en situasjon som gjør at de på en måte utvikler seg, og det er jo noe du gjør i forhold til et kompetansemål, men det er ikke noe du samtidig gir en veiledning om, ikke sant, men man har jo det i bakhodet.<sup>68</sup>

Samtidig, som det pekes på i artikkel 1, tyder resultatene på at det er grunn til å diskutere Black og Wiliams (2012a) antagelse om at feedback i kroppsøving, som eksempel på et fag preget av praktisk utøvelse, vil være mindre problematisk fordi hensikten med, og behovet for, feedback er tydelig for elevene og lærerne. Mot dette viser artikkel 1 at det ikke er tydelig for de fleste elever i avhandlingens seks skoler. Over halvparten av elevene

---

<sup>68</sup> Litt utvidet versjon av sitat som i artikkelen er oversatt til engelsk. Lærerens nummer i artikkelen er 2F, hvilket betyr den andre lærer som tok ordet i fokusgruppen ved skole F. Skole F tilsvarer skole nummer 6 i utvalgsbeskrivelsen i kappens kapittel 3.

rapporterte manglende feedback og bare ca. 15 % av elevene sa seg enig eller helt enig i at de får tilbakemeldinger om hvor de er i læreprosessen.

Feedback er spesielt interessant blant avhandlingens resultater fordi det er relatert til denne nøkkelstrategi for VfL at analysene avdekker et tydelig sprik mellom lærer- og elevperspektivet. I strid med at storparten av elevene i studien rapporterer få feedbackerfaringer virker kroppsøvlingslærerne ganske unisont å mene at de er gode på å gi tilbakemeldinger. Lærerne formidler en forståelse av undervisvurdering som noe de gjør kontinuerlig i kroppsøving. Dette skiller seg fra Vinje (2008a) som, i en undersøkelse blant kroppsøvlingslærere på ungdomsskoletrinn og i videregående skole i Oslo, fant at tre av ti av lærerne uttrykte at de grunnet liten tid ikke rakk å gi tilbakemeldinger til alle elever relatert til kompetansemål og delmål. Jeg vil kommentere at ulik metodetilnærming utgjør en mulig forklaring på forskjellen mellom undersøkelsene. Tallene fra den kvantitative undersøkelsen til Vinje (2008a), der enkeltlærere har svart ut fra egen individuelle praksis, kan være gyldige for avhandlingsstudiens 23 kroppsøvlingslærere sett under ett. Samtalen som gjennomføring av fokusgrupper innebærer, er også en reflekterende læringssituasjon for deltakerne (Kamberelis & Dimitriadis, 2013). Så selv om en forståelse av små tips og fremadrettede kommentarer i kroppsøvingstimene som undervisvurdering er noe som kommer til uttrykk ved alle seks skoler, er det sannsynlig at ikke alle deltakende lærere hadde bevisstgjort dette perspektivet før det kommer fram i løpet av fokusgruppesamtalen. Funn av variasjon i forståelse og utøvelse av vurdering mellom lærere ved samme skole støtter en slik fortolkning.

Det er flere positive trekk ved den kontinuerlige feedbackpraksis kroppsøvlingslærerne påpeker. Både Shute (2008) og Hattie og Timperleys (2007) oppsummerende analyser av tidligere feedbackforskning levner liten tvil om at umiddelbar feedback er å foretrekke. Det er sannsynlig at Black og Wiliams (2012a) argument i denne kontekst holder stikk, og at fagets praktiske egenart vil komme læreren (og den selvvaluerende elev) til gode ved at det i mange tilfeller er mulig å observere direkte om elevene har forstått og kan anvende informasjonen de får. Videre viste Kluger og DeNisis (1996) tjue år gamle omfattende og psykologisk orienterte review, hvordan feedback knyttet til enkle konkrete oppgaver var mye mer effektivt enn feedback knyttet til mer komplekse læringsoppgaver. Til det siste konkluderer de med at «the evidence for any learning effect here was minimal at best» (s. 278). På vektskålen for det som kanskje ikke er så positivt legger artikkel 1 hvordan kroppsøvlingslærernes omtale av små tips og konkrete ferdighetstilbakemeldinger som feedback på læringsmål kan utgjøre et bidrag til at kroppsøving i praksis reflekter det Kirk

(2010) beskrev som «physical education-as-sport-techniques» (s. 4). Det at feedback i lærernes beskrivelser virker å være knyttet til (å øve på og forbedre) ferdigheter som i stor grad har en lukket bevegelsesløsning definert av idrett og trening, medfører at «en bestemt type evne, bestemte ferdigheter og visse typer kropper verdsettes og favoriseres, mens andre marginaliseres og ekskluderes» (Svendby, 2013, s. 215). I motsetning til å støtte opp om et argumentert endringsbehov for å ivareta alle elevers læring og erfaring av meningsfull kroppsøving, vil denne praksis bidra til å opprettholde en idretts- og prestasjonsdiskurs som har preget kroppsøving også i Norge (Moe, Leirhaug & Resaland, 2015; Moen, 2011; Säfvenbom, Haugen & Bulie, 2015; se også artikkel 4).

Leahy m.fl. (2005) deler feedback og vurderingspraksis på klasseromsnivå inn i et kort, mellomlangt og langt tidsspenn<sup>69</sup>. Feedback i det korte tidsspennet foregår innenfor læringsøktene, dag for dag, minutt for minutt. Det mellomlange tidsspennet handler om hva som skal læres innenfor det emnet det arbeides med eller ut fra en periodeplan, mens feedback i det langsiktige tidsspennet knytter seg til målsettinger for et halvår eller år, inkludert et blikk enda lenger fram mot hvilken sluttkompetanse opplæringen i faget skal føre til. Det kroppsøvingslærerne i avhandlingsstudien beskriver som kontinuerlig praktisering av undervisvurdering må tolkes som feedback i det korte tidsspennet. De argumenterer selv for hvordan dette har en sammenheng med læreplanens kompetansemål og målsettinger i et lengre tidsspenn, men det blir samtidig tydelig at dette i liten grad er klargjort ovenfor elever, – og for den saks skyld: andre utenfor det som kan kalles kroppsøvingslærernes fagfellesskap.

I likhet med Shute (2008), som med utgangspunkt en review av 100 studier på feedback skriver at studiene inneholder «many conflicting findings and no consistent pattern of results» (s. 153), kan jeg ikke trekke noen entydige konklusjoner rundt kroppsøvingslærernes praktisering av framoverrettet feedback i lys av å være en VfL nøkkelstrategi. Det er derimot klart at den uttrykte forståelsen av feedback i det korte tidsspennet vil kreve en mer systematisk relatering til både langsiktige læreplanmål og feedback i det mellomlange og lange tidsspennet før denne praksis kan betraktes som valide undervisvurderinger. Videre er det, som lærerne selv er inne på, nødvendig med tydeliggjøring eller synliggjøring ovenfor elever (og andre) av feedbackpraksis i det korte tidsspennet som undervisvurdering nettopp i lys av de langsiktige perspektivene. Uten en klargjøring av disse sammenhengene er det vanskelig å se for seg hvordan lærerne skal kunne gi meningsfull og framoverrettet feedback på læring relatert til lengre tidsspenn, og i enda

---

<sup>69</sup> Jeg bruker her en norsk terminologi for vurderinger i ulike tidsspenn, slik de er presentert i boka *Feedback i skolen* (Hartberg, Dobson & Gran, 2012).

mindre grad hvordan de skal kunne skape vilkår der elevene gir feedback til hverandre eller henter tilbakemeldinger fra eget faglig arbeid og bevegelseserfaringer. I følge Hattie og Timperleys (2007) metaanalysestudie er det når feedback lykkes med å trigge elevenes metakognitive refleksjoner i læreprosessen at effekten blir mest virkningsfull.

Det at kroppsøvlingslærerne uttrykker å være tilfreds med den feedbackpraksis de beskriver kan føre til at de ikke reflekterer kritisk over det som kan utgjøre et validitetsproblem i forhold til læreplan og faglige målsetninger. Sammen med lite refleksjon over potensialet som ligger i å involvere elevene i større grad, kan dette avskjære dem fra å arbeide konstruktivt videre med kvalitet i feedback i ulike tidsspenn slik at det blir effektivt som VfL. Mangelen på uttrykk for gjennomtenkt dialog med elever og kritisk refleksjon over egen feedbackpraksis vitner om at kroppsøvlingslærerne har avgrenset forståelse av hvordan VfL strategien framoverrettet feedback skal fungere som et ledd i et lengre læringsforløp, og dette fører fint inn mot tematikk for neste avsnitt.

### **Elevinvolvering og kritisk refleksjon**

Resultatene relatert til elevenes perspektiv i artikkel 1 viser at det er VfL nøkkelstrategiene knyttet til egen- eller selv vurdering og medelev- eller hverandrevurdering som kommer desidert lavest ut. Dette harmonerer med hvordan lærerne uttrykker seg om det samme. I artikkel 3 og artikkel 4 kommer det fram at selv de av kroppsøvlingslærerne som har jobbet mest med egenvurdering blant sine elever beskriver dette til å være i en utforskende fase. I forhold til bruk av medelevvurdering uttrykker flere av lærerne at de er skeptiske og følgelig ikke har forsøkt det<sup>70</sup>.

Funn fra følgeforskningen for det nasjonale prosjektet *Bedre vurderingspraksis* forteller at det ikke bare er i kroppsøving lærere i liten grad involverer elevene i diskusjoner om vurderingsgrunnlaget (Stokke m.fl., 2008). Sammenholdt med resultatene i avhandlingen kan dette tyde på at implementeringen av LK06 og nye vurderingsformer i hovedsak har vært opptatt av formulering og tydeliggjøring av mål, kriterier og kjennetegn på måloppnåelse, mens det å utvikle strategier for reelt og verdifullt engasjement av elevene i vurderingsarbeid ser ut til å ha fått mindre plass. Nasjonale studier av feedback tyder på at lærere vurderer tilbakemeldingene til elevene som mer verdifull for elevene enn de er og konkluderer med at

---

<sup>70</sup> I lys av dette er det bemerkelsesverdig at 64 % av kroppsøvlingslærerne i Sandvik m.fl. (2012) sier «at de benytter medelevvurdering i noen eller stor grad» (s. 126). Studien som undersøker formativ vurdering i flere fag sier samtidig at kroppsøvlingslærerne sjeldnere enn de andre faglærerne bruker vurderingsresultater for å tilpasse egen undervisning og at de «benytter i mindre grad underveivurdering for å fremme elevenes videre læring» (s. 152).

det ikke er tilstrekkelig å informere elevene om mål og vurderingskriterier for å fremme læring (Havnes m.fl., 2012).

Internasjonalt er det påvist en rekke utfordringer med involvering av elever som sentrale strategier innen VfL. Black m.fl. (2003) noterte i rapporteringen fra deres VfL implementeringsstudie at lærernes første forsøk med egenvurdering blant elevene, og det å la elever sette egne læringsmål, ikke lyktes særlig godt. Elevene manglet både ferdighetene og kunnskapen til å kunne gjennomføre disse oppgavene på en meningsfull måte. I praksis, for at elevene skulle tilegne seg den nødvendige forståelse og kompetanse fant Black m.fl. (2003), at «peer-assessment turns out to be an important complement and may even be a prior requirement for self-assessment» (s. 50). De skriver helt konkret at utvikling av egenvurdering ble «approached through the development of peer-assessment practices in the teachers' classrooms», fordi dette hjalp elevene «to develop the objectivity required for effective self-assessment, which gave them both a concept of what quality meant in a specific context and also the motivation to seek improvement» (s. 31).

I lys av disse forskererfaringene er fomlingen blant kroppsøvingslærere som utforsker egenvurdering blant sine elever forståelig. Og kanskje er det bra at kroppsøvingslærerne har vært litt skeptiske og nølende til å starte medelevvurdering? Uten et solid forarbeid med å klargjøre kriterier og rammer for hva og hvordan det skal gis vurderinger, er det overhengende fare for at elevene stort sett vil kommentere god innsats og jobbing framfor å bidra med meningsfull læringsrettet feedback. Medelevvurdering forutsetter videre et trygt og åpent læringsmiljø, og et parallelt arbeid med å gjøre elevene vurderingsliterate, slik at de forstår for eksempel at medelevvurdering handler om å støtte læringsprosessen og ikke om å dømme en ferdig prestasjon. Når Ommundsen (2015) påpeker at et godt fysisk-motorisk og psykososialt læringsutbytte i kroppsøving handler om at lærere klarer å skape motivasjon, læringsberedskap, gode psykososiale læringsbetingelser og etablere positive følelser for kroppslig læring i og gjennom bevegelsesaktiviteter, er det en påminner om at VfL strategiene ikke er et magisk svar på de faglige utfordringer kroppsøving står ovenfor. Men som et ledd i det større læringsbildet kan de utgjøre et kraftfullt bidrag til å bevisstgjøre og innrette så vel fagdiskusjoner som praksis mot elevenes læring.

Med referanse til Ward og Lee (2005) påpekes i artikkel 3 at det sannsynligvis er et stort potensiale knyttet til å utvikle medelevvurdering. I deres review fremstilles hverandrevurdering som et relativt grundig utforsket område i kontekst av idrett og bevegelsesaktiviteter. Selv om mange av studiene handler om å utvikle vurderingsverktøy og teste reliabilitet og validitet i elevens vurderinger sammenlignet med ekspertvurderinger

knyttet til bestemte idrettsaktiviteter, demonstrerer studiene gjennomgående høy reliabilitet i elevenes vurderingsarbeid. Nadeau, Richard og Godbout (2008) er et eksempel på dette knyttet direkte til ishockey, mens Otero-Saborido og González-Jurado (2015) utvikler og tester et mer generelt verktøy for formativ vurdering i lagspill. Studier som disse forteller at medelever sannsynligvis kan bidra læringsstøttende og (lære seg å) gi framoverrettet feedback opp mot faglige mål. Men ved bruk av slike vurderingsverktøy inn i kroppsøvingssammenheng må det følge en vurdering av validitet relatert til skolesystem og kroppsøvingssammenheng. Læringsplanmål, noe som igjen understreker det gjennomgående behovet for kritisk refleksjon i arbeidet med vurdering.

I tillegg til utfordringene med lav grad av elevinvolvering framtrer i artikkel 3 det jeg drister meg til å kalle et skremmende fravær av kritisk refleksjon relatert til vurdering. Dette er hovedbidraget til utforskningen av «assessment literacy» blant kroppsøvingslærerne og viser verdien av å bruke teori som analyseverktøy. Elementet «critical engagement with assessment» hos Hay og Penney (2013) gir mulighet til å løfte fram poeng som er fraværende i det empiriske materialet og derfor ville vært vanskelig tilgjengelig gjennom en tematisk analyse hvor induktiv tilnærming rendyrkes. Denne erfaringen yter indirekte støtte til å styrke teoriens rolle i arbeid med profesjonell utvikling og å styrke kontinuerlig kritisk refleksjon hos kroppsøvingslærere og i kroppsøvingslærerutdanningene. Referanser til læringsteori og annen bruk av teori som bakgrunn for refleksjon og undervisningsplanlegging har vært omstritt og lite framtrekkende i kroppsøvingslæreres (og kroppsøvingslærerutdanningenes) praksis (Kjerland, 2015; Moen & Green, 2014). Hay og Penney (2013) skriver at den vurderingsliterate lærer er klar over at ingen vurderingspraksis er nøytral og at vurdering i skolen alltid ledsages av en uungåelig og assymetrisk maktdistribusjon. Vurderingshandlinger er bærere av mer eller mindre eksplisitte verdiladninger og virker i sosiale settinger eller felt der de vil ha både tilsiktede og utilsiktede konsekvenser. Å være oppmerksom på dette, hvilket kroppsøvingslærerne i studien ikke er, omtaler Hay og Penney (2013) som lærernes refleksive kapasitet, noe som vil inkludere selvanalyse av hvordan egen historie, egne verdier og fordommer, influerer på oppfatningen av og interaksjoner med elever.

En sammenligning av resultatene i artikkel 3 og Dinan Thompsen og Penneys (2015) studie av «assessment literacy» blant lærere i «primary physical education» indikerer at kroppsøvingslæreres utfordringer relatert til «assessment literacy» krysser både landegrensene og elevgrupper i ulike aldre. Med tanke på endring i vurderingspraksis og implementering av VFL strategier virker det riktig å understreke ett poeng hos Engvik (2013): «Elevinvolvering krever endringer som læreren kan gjøre i egne timer, og er slik sett ikke avhengig av tett



kollegialt samarbeid» (s. 98). Ved godt planlagt og gjennomført elevinvolvering kan elevgruppen utgjøre en diskusjonspartner og rettesnor for læreren også der det ikke (ennå) er rom for moderering av vurderingspraksis i fagfellesskap med kolleger<sup>71</sup>. Det er en gjensidighet i dette. I tråd med Hay og Penneys (2013) forespeiling vil økt elevinvolvering kreve at elevene må gjøres vurderingsliterate. For at elevene skal kunne bli vurderingsliterate, eller reelle deltakere i vurderingsdiskursen ved egen skole, må lærerne selv være literate i vurdering. Det er en forutsetning, som tidligere nevnt, at lærerne er godt skolerte i vurdering for at et kriterie-referensielt vurderingssystem skal kunne fungere mest mulig rettferdig. Ved å planlegge for å involvere elevene i større grad må de også ha forståelse for hvordan dette innebærer en forskyvning i maktrelasjoner og kan kreve større endringer i undervisningen (se Tolgfors & Öhman, 2015; Palm, 2014). Økt elevinvolvering vil kunne<sup>72</sup> demokratisere vurderingspraksis på måter som både er en forutsetning for god og velintegret VfL og være et steg mot å sikre gjennomsiktighet som ledd i å gjøre vurdering mer rettferdig (og valid) i et skolesystem der ansvaret for faglig vurdering i kroppsøving er desentralisert, mange steder i praksis ned til den individuelle lærer.

### **VfL i kroppsøving i Norge nå og i fremtiden**

En revidert læreplan i kroppsøving med tilhørende justering av vurderingsregelverket ble iverksatt med virkning fra skoleåret 2012-2013 (Utdanningsdirektoratet, 2012)<sup>73</sup>. Ny læreplan nesten umiddelbart etter at datamaterialet til denne avhandlingen var samlet gjør spørsmålet påtrengende: Hvilken betydning kan det ha for resultatene? Mitt svar er at jeg mener det har liten betydning til tross for at det ble gjort endringer som direkte berørte vurderingsgrunnlaget i faget kroppsøving. Sammenheng mellom fagets formål og kompetansemålene skulle tydeliggjøres, elevens innsats kom inn som del av vurderingsgrunnlaget og et tidligere skille mellom grunnskole og videregående skole i

---

<sup>71</sup> Vurderingshistoriene til de tre lærerne Geir, Leif og Ronald i artikkel 4 eksemplifiserer alle Engviks (2013) poeng, og mulighetene som kan ligge i økt elevinvolvering er i artikkel 1 spesielt løftet fram: «It may be that the students' voices about their experiences in PE classrooms create a space where teachers and students can participate in a dialogue about what is relevant and worthwhile learning in PE. Moreover, they could together explore how to elicit evidence of such learning, as well as how to use that evidence to make decisions about the next step in the instruction and learning processes» (Leirhaug & Annerstedt, 2015, s. 13).

<sup>72</sup> Formuleringen «vil kunne» er her som en del andre steder helt presis. Avhandlingen diskuterer jo det med elevinvolvering på grunn av at analysene avslører bruk av egen- og medelevvurdering som sjeldent eller fraværende i den vurderingspraksis som lærerne beskriver, dermed er det ikke grunnlag for å si mer enn det normative «vil kunne» og støtte denne pedagogiske hypotese på de få småskalastudiene som finnes i kroppsøving (for eksempel Ní Chróinín & Cosgrave, 2013; Hofnes, 1997, 2000).

<sup>73</sup> Det er også gjort endringer i læreplan for kroppsøving etter dette. Kompetansemål i svømming og livredning ble betydelig endret med virkning fra august 2015 (Utdanningsdirektoratet, 2015b). Relatert til VfL og de poenger som diskuteres i avhandlingen er endringene etter min vurdering ikke betydningskillende.

forhold til elevens forutsetninger ble fjernet, men hva gjelder VfL og krav til formative vurderingspraksiser så innebar revisjonen en kontinuering heller enn endring og brudd. I klartekst betyr det at de utfordringene og innsatsområdene som er demonstrert for kroppsøving gjennom avhandlingen består. Dette gjenspeiles i Kunnskapsdepartementets uttrykte intensjon om å videreføre den nasjonale satsingen på VfL som ledd i kompetanseutviklingen blant lærere (Meld St. nr. 20 (2012-2013), 2013), og igjen finnes i innstillingen om framtidens skole, hvor Ludvigsen-utvalget understreker at underveisvurdering «er en integrert del av en læringsfremmende undervisningspraksis i fagene, og bør prioriteres i kompetanseutviklingen» (NOU 2015:8, 2015, s. 85).

Det ble også gjort justeringer i det generelle vurderingsregulerende lovverket fra høsten 2015 (Utdanningsdirektoratet, 2015a). To momenter<sup>74</sup> i denne justeringen berører resultatene i avhandlingen på en måte som kan være avgjørende. For det første er det gjort en endring som forsterker og tydeliggjør sammenhengen mellom underveisvurdering og sluttvurdering i fag. Dette er en endring som støtter opp under de utfordringer som er diskutert på bakgrunn av manglende sammenheng mellom VfL og elevenes karakterer i kroppsøving i artikkel 2. Etter forskriftsendringen sies det eksplisitt at den kompetanse eleven har vist underveis i opplæringa skal være en del av grunnlaget for standpunktarakter i fag. Det andre momentet er fjerningen av kravet til dokumentasjon av underveisvurdering. I kroppsøvingsteksten kan dette kravet se ut til å ha styrket argumentasjon for bruk av tradisjonelle tester knyttet til vurdering av ferdighet og fysisk styrke og form (Elnan & Sando, 2014; Rønninghaug, 2011). Læreren Leif i artikkel 4 er en av de lærerne som sier at de gjør en del som har med vurdering å gjøre mest for å krysse av at de har gjort det. Slik sett faller de inn i den kategorien Wilson (2014) viste at kunne oppfylle profesjonskravene fra utdanningsmyndighetene uten å drive formativ vurdering slik forskningen på formativ vurdering presenterte den. Dette (Leifs praksis) blir da et uttrykk for det som i artikkelen kalles vurdering for vurderingens skyld, og ikke VfL. Dokumentasjonskravet har kanskje vært en pådriver til at lærere ved alle studiens skoler arbeider med å endre vurderingspraksis, men det har muligens på samme tid skapt hindringer for utvikling og implementering av mer genuine VfL strategier i kroppsøving.

Støttet av tidligere forskning i Norge på vurdering i kroppsøving etter LK06 (Arnesen, Nilsen & Leirhaug, 2013; Prøitz & Borgen, 2010; Sandvik & Buland, 2014; Vinje 2008a), viser foreliggende avhandling at VfL ikke er implementert i kroppsøving (i de seks

---

<sup>74</sup> Begge momenter er resultat av endring i §3-16 i Forskrift til opplæringslova (2006).

videregående skolene) på en måte som gjør det mulig å trekke konklusjoner om hvordan VfL vil fungere og slå ut i kroppsøving, hvilke betydninger det faktisk kan ha for elevene og læring i faget. Av dette følger også at vi ennå ikke har tilstrekkelig kunnskap om betydning og konsekvenser av VfL som del av et mer systematisk vurderingsarbeid i kroppsøving. Denne kunnskap finnes heller ikke internasjonalt (Georgakis & Wilson, 2012; López-Pastor m.fl., 2013). Avhandlingen har bidratt til å peke ut sannsynlig viktige innsatsområder i det videre arbeidet, både i Norge og i andre land. Utfordringene ved implementering av VfL er litt ulike, men det overordnede utfordringsbildet, som er gjenkjennbart i denne avhandlingens resultater, går på tvers av utdanningssystemene i den vestlige verden (Berry & Adamson, 2011; Birenbaum m.fl., 2015). Et område som ikke er berørt i avhandlingen, men som internasjonalt har fått begynnende fokus i sammenheng med vurdering i kroppsøving, er utvikling og utprøving av digitale vurderingsverktøy (se for eksempel Penney m.fl., 2012; Weir & Connor, 2009). Forskning og utvikling av nye digitale læringsplattformer og vurderingsverktøy vil sannsynlig bidra til andre måter å tenke og tilrettelegge for framoverrettet feedback og egenvurdering, så vel som mulighet til å følge elevers læring og utvikling over tid i kroppsøving. En slik utvikling vil ha betydning for forståelse og implementering av VfL i kroppsøving og understøttes av NOU 2015:8 (2015) om fremtidens skole der undervisvurdering løftes opp som «et av områdene der digitalisering kan bidra til å støtte elevene og lærerne på andre måter enn det som er praksis i dag» (s. 81).

Spenninger mellom formative og summative vurderingspraksiser, VfL og undervisvurdering, digitalt eller ikke? Det understrekes i NOU 2015:8 (2015) om fremtidens skole at undervisvurdering er et viktig virkemiddel for å fremme elevenes læring og presiseres at det «bør vektlegges som en integrert del av undervisningspraksisen i fagene» (s. 13). Avhandlingens resultater indikerer at mange norske kroppsøvingslærere vil trenge både tid og substansiell støtte for å opparbeide den kunnskap og kompetanse de vil trenge for å kunne realisere slike visjoner. Vesentlige satsingsområder som gjennom avhandlingens artikler er identifisert med tanke på VfL i kroppsøving er (i) å utvikle lærernes forståelse av VfL og dens potensielle betydning i opplæringen, med et spesielt blikk på (ii) deres kritiske holdning til og refleksjoner over utøvd vurdering. Det siste berører både forståelse av hvordan elevers bakgrunn, kultur og sosiale forskjeller kan innvirke på vurderinger, og et behov for (iii) å utvikle et felles fagdidaktisk begrepsapparat eller språk egnet for vurdering i kroppsøving. Videre er det, slik det også innledningsvis ble lagt fram med referanse til Georgakis og Wilson (2012), et stort behov for (iv) å utvikle og beskrive kroppsøvingsspesifikke eksempler på VfL som en del av fagdidaktikken, samt utforske

sammenhengen mellom slike praksiser og elevenes læringsutbytte. Til slutt, og sannsynligvis integrert i arbeid med de andre innsatsområdene, er (v) et gjennomgående behov for å heve elevenes forståelse av og deltakelse i vurdering på plan- så vel som prosessnivå.

Validitet og sosial rettferdighet vil hele veien være fundamentale orienteringspunkter i dette arbeidet (Hay & Penney, 2013; Klenowski & Wyatt-Smith, 2014), og avhandlingens resultater antyder at en utvikling av begrepet «assessment literacy», eller den vurderingsliterate lærer, kan gi en fruktbar ramme til arbeidet. Uansett tilnærming, vil det for å følge opp konsekvenser for elever og sikre validitet, være nødvendig med forsknings- og utviklingsarbeid. Til bildet hører også at lærerutdanningene har et ansvar for å besørge oppdatert og god undervisning om vurdering slik at nyutdannede kroppsøvingslærere kan fungere som en ressurs for vurderingspraksis i skolene. Lorente-Catalán og Kirks (2016) studie av en VfL undervisningsmodul i kroppsøvingslærerutdanning tyder på at studenter som møter det i utdanningen opparbeider tro på betydningen av VfL i kroppsøving og uttrykker at de er motiverte for å ta det i bruk i egen framtidig lærerpraksis.

For å avslutte er det fristende å gripe tilbake til begynnelsen av kappens kapittel 2 og den andre grunngevingen for formativ vurdering som har blitt hengende ved meg. Forstått som innflettet eller integrert i selve læreprosessen vil formative vurderingspraksiser sannsynligvis – enten de kalles VfL eller går under annet navn – være en del av god og effektiv pedagogikk. Selv ikke kritikere som Bennett (2011) og Torrance (2012) virker å tvile på effekten hvis og når VfL er velintegrert i læreres og elevers hverdagspraksis, fagdidaktisk tilpasset og forankret i læreplan og læringsteori. I tilknytning til slike forutsetninger for vellykket VfL implementering stilles i artiklene spørsmål både ved at det hersker lite klarhet og enighet om det som kan kalles kunnskapssubjektet, eller gyldig og verdsatt kunnskap, i kroppsøving (Annerstedt, 2010; Dyson 2014; Hay, 2006), og (påviste utfordringer) ved at læringsteoriene som har dominert i pedagogikken i all hovedsak har vært kognitivt orientert, mens kroppsøving i stor grad er preget av bevegelse og kroppslig læring (Quennerstedt, Öhman & Armour, 2014; Stoltz, 2014). Spørsmål som dette har kommet opp i randsonen av studiens problemstillinger, da vurdering som læringsfremmende verktøy med nødvendighet er tett knyttet til spørsmålene om hva som skal læres og hvorfor. Å utdype, forfølge og eventuelt besvare slike spørsmål har ikke vært en del av denne avhandlingens oppgave, men de peker mot at det ennå er mye forskning og kunnskapsproduksjon som venter i og rundt kroppsøving i skolen.

## Kapittel 6. En empirisk epilog

Et sitat fra Ronald innleitet avhandlingen. Ved innlevering står Ronald på terskelen til pensjon. Før skoleslutt våren 2015 gjennomførte jeg et oppfølgingsintervju med Ronald der hensikten blant annet var å undersøke hvordan hans tanker og praksis rundt VfL i kroppsøving hadde utviklet seg. Fullt klar over at det ikke er vanlig praksis å innskrive ny empiri og resultat i kappen, har jeg innvilget meg et unntak i denne epilogen. Siden Ronald ble tildelt æren av å åpne avhandlingen ble det fristende også å la hans stemme avrunde.

Intervjuet fant sted i et møterom på skolen der Ronald jobber og varte i 48 minutter. Det ble gjennomført ut fra informert samtykke og var et semistrukturert kvalitativt intervju<sup>75</sup>. En intervjuguide var utarbeidet med utgangspunkt i temaguiden for fokusgruppene (vedlegg 4), analyser av fokusgruppene, Ronalds vurderingshistorie slik den er konstruert i avhandlingens fjerde artikkel (Leirhaug & MacPhail, 2015), og noen andre spørsmål omkring det å være heltids kroppsøvingslærer fram mot pensjonsalder (gitt at mange slutter eller reduserer sin stilling tidligere i kroppsøving sammenlignet med lærere i andre fag; se Jakobsen, 2010). Intervjuet inkluderte også en del der han ble invitert til å fortelle om sitt syn på kroppsøvingsfagets utvikling og å spekulere over fagets potensielle framtider. Som grunnlag for epilogen benyttes kun de deler av intervjuet med relevans for avhandlingens resultater og VfL som tema.

Jeg vil starte med hvordan Ronald peker på innføringa av LK06 som noe som virkelig har endret seg i perioden han har vært kroppsøvingslærer:

Ronald: Kunnskapsløftet i 2006, som ikke da førte med seg så mye endringer, det har jo tatt ganske lang tid. Jeg vil jo egentlig kanskje si at det tok seks år før det begynte å skje no.

Forsker: akkurat, ja...

Ronald: Sånn er min opplevelse av det.

---

<sup>75</sup> I stedet for å oppgi nye referanser til metodelitteratur over hvilke utfordringer som kan forventes i planlegging, gjennomføring, analyse og rapportering av et slikt kvalitativt intervju, noe de fleste større bøker om kvalitativ metode og intervju kan bidra med, så mener jeg det er tilstrekkelig å vise til de metodebøkene som allerede er benyttet i kappens kapittel 3, for eksempel bidrar Bryman (2012) med en kort, men informativ, innføring i kvalitative intervju. Det sagt, vil jeg tillegge at jeg i oppstarten av mitt doktorgradsarbeid ble ganske inspirert av å lese *Interpreting interviews* av Alvesson (2011).

At han løfter fram innføringa av LK06 kan både være influert av at det er siste læreplan og koples til at endringene i vurderingsregelverket ved denne reformen ble opplevd spesielt store for kroppsøving (Engvik, 2010; Leirhaug, 2016; Vinje & Brattenborg, 2016). Det virker likevel å være mer generelle og dyptgripende endringer Ronald sikter til. På spørsmål om hva endringa besto i, tenker han seg om og sier:

Endringa er vel at en har begynt å se hva kompetansemålene skulle være, hva kjennetegn på måloppnåelse skulle være, at det begynner å bli brukt og at det kommer i et system. At en begynner å forholde seg til det på en måte sånn at det også kommer ut til elevene.

Det siste er viktig relatert til resultatene i avhandlingen. Hvis Ronalds beskrivelse er noenlunde dekkende, at det omtrent tok seks år før implementeringen av LK06 begynte å fungere for elever i kroppsøving, så vil det kunne forklare deler av avhandlingsstudiets resultat der elever rapporterer få VFL erfaringer til tross for at flere lærere var opptatt av det. Tidsperspektivet støttes av andre resultater i avhandlingen. For eksempel kommer det fram i avhandlingens tredje artikkel (Leirhaug, MacPhail & Annerstedt, 2016) at matrisene for vurdering i enkeltaktiviteter som lærerne på en skole tilsynelatende hadde brukt mye tid på å utarbeide, også som ledd i et større nettverksarbeid på tvers av skoler i fylket, var noe de ennå ikke hadde presentert for eller brukt sammen med elevene.

Ronald representerte i 2011/2012 en av avhandlingsstudiets VFL interesserte og oppdaterte lærere, og beskrev allerede da hvordan han hadde endret en god del i sitt vurderingsarbeid. Ronald er nå som da opptatt av målene i faget og hva elevene skal lære. Han snakker om planlegging, hvordan han kan «ta utgangspunkt i et kompetansemål, velge en aktivitet og lage et opplegg for det, og så beskrive kjennetegn på måloppnåelse». Ronald kommer ikke med eksempler på, og virker heller ikke å vurdere, for øvrig i likhet med et gjennomgående trekk blant lærerne i avhandlingsstudien, å inkludere elevene i arbeidet med å utforme kjennetegn på måloppnåelse. Jeg spør mer direkte om det med inkludering av elever i vurderingsarbeidet. Han svarer å ha

...prøvd litt variabelt, det er også noe med at det kanskje er litt tidkrevende. Og så er det noe med at man er litt aleine, ikke sant? Det er jo ikke noen oppskrift på detta, så du må prøve deg ut. Jeg prøvde meg ganske mye til å begynne med, men så ser jeg at det blir forferdelig mye papirarbeid for meg og hva får jeg egentlig igjen? Og hva får

elevene igjen? Så jeg har redusert det betydelig, men jeg har kommet fram til noen sånne hensiktsmessige ting. Som for eksempel at de [elevene] i grupper kan ha et opplegg med framføring og at de etter framføringa har en samtale hvor de nærmest har en sånn «two stars and a wish».

Da Ronald jobber ved en større skole og har flere kolleger som har full stilling som kroppsøvingslærere er det bekymringsfullt å høre hvordan han beskriver at han opplever å være alene. Teori forteller at arbeid og moderering i fagdidaktisk fellesskap med andre lærere er fundamentalt både for validitet i læringsorientert vurdering og for profesjonell læring (Engvik, 2013; Klenowski & Wyatt-Smith, 2014). Det kan virke som det å utvikle strategier for bedre vurderingspraksis i første rekke blir relatert hva enkeltlæreren gjør i møtet med egne elever, og i mindre grad fører til faglig utviklingsarbeid og utadrettet virksomhet som vil bidra til å styrke profesjonaliteten for kroppsøvingslærerne mer generelt. Hva får han igjen? Hva får elevene igjen? Spørsmålene forteller at Ronald slett ikke er overbevist om at en del av det han har prøvd ut er VfL. Uten at han selv bruker begrepet validitet, setter han indirekte spørsmålstegn ved validiteten i noen av de formative vurderinger han har iverksatt. Sammen med poengtering av tiden det koster, tydeliggjør dette at Ronald så langt ikke har noen entydig erfaring av verdien av å engasjere elevene for eksempel gjennom egenvurdering. Bildet nyanseres ved at han viser til enkelte verktøy som han har funnet det hensiktsmessig å videreføre og gjort til sine egne<sup>76</sup>. Å kunne fått delt og reflektert over disse erfaringene i et vurderings- og tolkningsfellesskap med andre lærere ville muligens kunne verifisere erfaringene, gjøre dem mer eksplisitte og videreutvikle foreliggende positive erfaringer til mer generelle kroppsøvingsdidaktiske strategier. Samtidig er det å benytte andres verktøy for å implementere VfL i egen undervisning sjelden en suksess. «Given that each individual teacher has to make the ideas their own, diversity is bound to go with success» (Black m.fl., 2003, s. 119).

Ronald er helt klar på hvordan LK06 har bidratt til at han arbeider mer systematisk og bedre med vurdering. Han forteller også hvordan han etter hvert ikke er helt fortrolig med å bruke begrepet VfL. Han reflekterer og sier at «vurdering for læring skal være noe du har hele tida, som en del av det du gjør». Da er det ikke så viktig om det heter VfL eller noe annet. Ronald vil like gjerne snakke om «undervisning som fører til læring», som VfL. Han konkretiserer først ved å si at å «snakke med eleven er viktigere enn det jeg skriver på

---

<sup>76</sup> Ronald viste til dette med «two stars and a wish» også under fokusgruppeintervjuet i 2012. Se Leirhaug og MacPhail (2015).

Fronter», og som følgende sitat viser, griper han til forklaringer i tråd med feedback-perspektivene som kom fram i avhandlingens studier når han skal utdype:

Ronald: Det er noe med vurdering for læring. Jeg tenker på vurdering for læring; den viktigste vurderinga jeg gjør er jo i timen.

Forsker: Ja?

Ronald: ...hvor jeg sier til eleven: «Kan ikke du gjør sånn?» Ikke sånn altså, men forklare hvordan et grunnskudd eller lay-up i basket skal være. Det er jo den viktigste vurderinga for læring, ikke sant? I hvert fall i forhold til den enkelte aktivitet. Mens mere på det som går på innsats, deltakelse, relasjoner, går kanskje mere på en sånn samtale?

Ronalds utvikling og praksis i forhold til VfL virker å være preget av at han var motivert men manglet midler. Han har fortsatt tro på at VfL har noe å bidra med til kroppsøving, men uttrykker samtidig erfaringer av at det er arbeidskrevende, noe som ganske sikkert er forsterket av at han har følt seg litt alene med å iverksette VfL nøkkelstrategier og at endring er et langtidsprosjekt. Motivasjonen for å prøve ut nye ting er ikke like stor lengre. Dette trenger ikke bety at han gjør mindre, det kan dels også skyldes at VfL «usually is informal, embedded in all aspects of teaching and learning, and conducted by different teachers as part of their own diverse and individual teaching styles» (Black m.fl., 2003, s. 2).

Ronald sier rett ut om VfL at det «er jo på en måte et slags nytt begrep som har kommet. Læreren har jo alltid gjort det, ikke sant, uten at det begrepet har vært der.» Dette er en innsikt han deler med Harry Black (1986), som i kapitlet med tittel *Assessment for learning* for tretti år siden introduserte emnet ved å peke på at formativ vurdering alltid har vært en del av læreres praksis. For å underbygge påstanden viser han til et brev<sup>77</sup> fra det 19. århundre skrevet av rektor ved Greenwich Hospital School. I lys av kroppsøvingsfagets vurderingshistorie er det mulig at Harry Black og Ronald tar munnen for full. Det er ikke sikkert formativ vurdering, og i sterkere grad VfL, hadde noen sentral rolle å spille i kroppsøving preget av drill og linjegymnastikk og etter hvert skoleidrett i etterkrigsåra. Det er ikke usannsynlig at oppførsel, gjennomføring, aktivitet og prestasjon har vært minst like framtrendende som læring. Uten å problematisere dette ytterligere, er det fortsatt ubesvart hva Ronald sikter til. Det må handle om mer enn forklaringer og tips til «grunnskudd eller lay-up

---

<sup>77</sup> Det er sitert etter E.B. Chadwick (1864): Statistics of educational results [Museum: a Quarterly Magazine of Education, Literature and Science 3, 479-484]. Jeg har ikke lest eller sjekket kilden.



i basket». Det gjenstår å finne ut hva disse «gode lærere» alltid har gjort, og om det kan utvikles, formidles og gjøres på måter som sikrer at det i utgangspunktet skal bidra til alle elevers læring og utvikling i og gjennom kroppsøving.



## Referanser

- Aasen, P. (2012). Accountability under ambiguity. Dilemmas and contradictions in education. I Østern, A.L., Ryghaug, T., Smith, K., Krüger, T. & Postholm, M.B. (red.). *Teacher Education Research between National Identity and Global Trends* (s. 77-92). Trondheim: Akademika.
- Aasen, P., Møller, J., Rye, E., Ottesen, E., Prøitz, T.S. & Hertzberg, F. (2012). *Kunnskapsløftet som styringsreform – et løft eller et løfte*. Rapport 20/2012. Oslo: NIFU. Hentet 12.6.2015 fra <http://www.udir.no/Tilstand/Forskning/Rapporter/NIFU/Kunnskapsloftet-som-styringsreform/>
- Alvesson, M. (2011). *Interpreting interviews*. London: Sage.
- Alvesson, M. & Sköldbberg, K. (2009). *Reflexive Methodology: new vistas for qualitative research (2<sup>nd</sup> edn)*. London: Sage.
- Annerstedt, C. (2007). *Att (lära sig) vara lärare i idrott och hälsa*. Göteborg: Multicare AB.
- Annerstedt, C. (2010). Karaktersetting i kroppsøving: Problematisk, urettferdig og neppe likeverdig I Steinsholt, K. & Gurholt, K.P. (red.). *Aktive liv: Idrettspedagogiske perspektiver på kropp, bevegelse og dannelse* (s. 233-254). Trondheim: Tapir.
- Annerstedt, C. & Larsson, S. (2010). 'I have my own picture of what the demands are...?' Grading in Swedish PEH – problems of validity, comparability and fairness. *European Physical Education Review* 16(2), 97-115.
- Aristoteles. (1999). *Den nikomakiske etikk*. Oslo: Gyldendal; De norske bokklubbene.
- Arnesen, T.E., Blair, B., Engelsen, K.S., Gamlem, S.M., Helleve, I., Olsen, J.I., Sandal, A.K. & Smith, K. (2010). *Kartlegging av vurdering som tema innenfor lærerutdanningen i UH-nett Vest. Samarbeidsrapport mellom universitet og høyskolene innenfor UH-nett Vest*. Rapport UH-nett Vest. Bergen: HIB, HISF, HSH, HVO, UiB.
- Arnesen, T.E., Leirhaug, P.E. & Nilsen A.K. (2010). Kristin Halvorsen og vi om innsats, testing og karaktervurdering. *Kroppsøving* 60(6), 22-23.
- Arnesen, T.E., Nilsen A.K. & Leirhaug, P.E. (2013). «Den læreplanen som ikkje kan tilpassast mi undervisning, finst ikkje.» Vurdering og undervisning i kroppsøving etter Kunnskapsløftet. *Tidsskriftet FOU i praksis* 7(3), 9-32.
- Augustad, P. (2003). *Skolering av kroppen. Om kunnskap og makt i kroppsøvingsfaget*. Doktorgradsavhandling, UiO. Bø: Høgskolen i Telemark.
- Bach, E. (2015). *Kunsten å vurdere elever i kroppsøving. En kvalitativ studie av åtte læreres vurderingspraksis etter innføringen av den reviderte læreplanen i 2012*. Masteroppgave. Oslo: Norges idrettshøgskole.
- Backman, E. & Larsson, L. (red.) (2013). *I takt med tiden? Perspektiv på idrettslærerutbildningen i Skandinavien*. Lund: Studentlitteratur.
- Bailey, R., Armour, K., Kirk, D., Jess, M., Pickup, I. & Sandford, R. (2009). The educational benefits claimed for physical education and school sport: an academic review. *Research papers in education* 24(1), 1-27.
- Baird, J.-A., T.N. Hopfenbeck, T.N., Newton, P., Stobart G. & Steen-Utheim, A.T. (2014). *State of the Field Review. Assessment and Learning*. Rapport 13/4697. Oslo: Norsk kunnskapssenter for utdanning.
- Barone, T. (1995). Persuasive writings, vigilant readings, and reconstructed characters: the paradox of trust in educational storysharing. *International Journal of Qualitative Studies in Education* 8(1), 63-74.

- Bennett, R.E. (2011). Formative assessment: a critical review. *Assessment in education: Principles, Policy and Practice* 18(1), 5-25.
- Berger, P.L. & Luckmann, T. (1966). *The social construction of reality. A treatise in the sociology of knowledge*. New York: Anchor books.
- Bernstein, B. (1975). *Class, codes and control, volume 3: Towards a theory of educational transmission*. London: Routledge and Kegan Paul.
- Berry, R. (2011). Assessment reforms around the world. I Berry, B. & Adamson, B. (red.). *Assessment Reform in Education: Policy and Practice* (s. 89–102). Dordrecht, Nederland: Springer.
- Berry, R & Adamson, B. (red.) (2011). *Assessment Reform in Education: Policy and Practice*. Dordrecht, Nederland: Springer.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education* 2(3), 347-364.
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L. & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education* 13(1), 117-140, DOI:10.1177/1478210314566733.
- Black, H. (1986). Assessment for learning. I Nuttall, D.L. (red.). *Assessing educational achievement* (s. 7-18). London: Falmer Press.
- Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice* 5(1), 7-73.
- Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80(2), 139-148.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. I Gardner, J. (red.). *Assessment and learning* (s. 81-100). London: Sage.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Education, Assessment, Evaluation and Accountability* 21, 5-31.
- Black, P. & Wiliam, D. (2012a). Developing a theory of formative assessment. I Gardner, J. (red.). *Assessment and Learning (2<sup>nd</sup> edn)* (s. 206-229). London: Sage.
- Black, P. & Wiliam, D. (2012b). The reliability of assessments. I Gardner, J. (red.). *Assessment and Learning (2<sup>nd</sup> edn)* (s. 243-263). London: Sage.
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning. Putting it into practice*. Maidenhead: Open University Press.
- Bloom, B.S., Hastings, J.T. & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: MacGraw-Hill.
- Bomo, S.Ø. (2008). *Implementering av Kunnskapsløftet*. Masteroppgave. Oslo: Norges idrettshøgskole.
- Borgen, J.S. & Engelsrud, G. (2015). Hva skjer i kroppsøvingfaget? *Bedre skole*, nr. 2, 62-67.
- Brattenborg, S. (1995). *Karaktersetting i kroppsøving på ungdomsskoletrinnet*. Hovedfagsoppgave. Levanger: Høgskolen i Nord-Trøndelag.
- Brattenborg, S. & Engebretsen, B. (2007). *Innføring i kroppsøvingdidaktikk*. Kristiansand: Høgskoleforlaget.
- Brattenborg, S. & Engebretsen, B. (2013). *Innføring i kroppsøvingdidaktikk*. Oslo: Cappelen Damm
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

- Brookhart, S.M. (2007). Developing measurement theory for classroom assessment purposes and uses. I McMillan, J.H. (red.). *Formative classroom assessment, theory into practice* (s. 43–62): New York: Teachers College Press.
- Bryman, A. (2008). *Social research methods* (3<sup>rd</sup> edn). Oxford, New York: Oxford University Press.
- Bryman, A. (2012). *Social research methods* (4<sup>th</sup> edn). Oxford, New York: Oxford University Press.
- By, I.-Å. (1998). *Utvikling av læreplanen i kroppsøving i grunnskolen i perioden 1922-1997*. Notat. Oslo: Norges idrettshøgskole.
- By, I.Å., m.fl. (2008). *Veiledning for vurdering med karakter i kroppsøving i kunnskapsløftet - videregående trinn 1, 2 og 3*. Oslo: LFF/Norges idrettshøgskole.
- Carroll, B. (1994). *Assessment in Physical Education: A Teacher's Guide to the Issues*. London: Falmer.
- Casey, A. (2013). Practitioner research: A means of coping with the systemic demands for continual professional development? *European Physical Education Review*, 19(1), 76-90.
- Chan, K., Hay, P. & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in Physical Education. *Asia-Pacific Journal of Health, Sport and Physical Education* 2(1), 3-19.
- Chatzopoulos, D., Erdmann, R. & Tsormbatzoudis, H. (2006). Pupils' grading: Do teachers grade according to the way they report? *International Journal of Physical Education* 43(1), 4-10.
- Chen, W. (2005). Examination of curricula, teaching practices, and assessment through national standards. *Physical Education and Sport Pedagogy* 10(2), 159-180.
- Creswell, J.W. (2009). *Research design. Qualitative, Quantitative, and Mixed methods approaches*. London: Sage.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research* 58(4), 438-481.
- Curtner-Smith, M.D. (1999). The more things change the more they stay the same: Factors influencing teachers' interpretations and delivery of national curriculum physical education. *Sport, Education and Society* 4(1), 75-97.
- Dalen, T. & Aune, T.K. (2013). Relativ alderseffekt ved karaktersetting i skolen. I Pareliussen, I., Moen, B.B., Reinertsen A. & Solhaug, T. (red.). *FoU i praksis 2012 conference proceedings* (s. 62-68). Trondheim: Akademika forlag.
- Dale, E.L. (2008). *Fellesskolen – reproduksjon av sosial ulikhet*. Oslo: Cappelen Akademisk.
- Dale, E.L. (2010). *Kunnskapsløftet – på vei mot felles kvalitetsansvar*. Oslo: Universitetsforlaget.
- Dale, E.L. & Wærness, J.I. (2006). *Vurdering og læring i en elevaktiv skole*. Oslo: Universitetsforlaget.
- Davis, A. (2009). Examples as Method? My attempts to Understand Assessment and Fairness (in the spirit of the Later Wittgenstein). *Journal of Philosophy of Education* 43(3), 371-389.
- Dinan Thompson, M. & Penney, D. (2015). Assessment literacy in primary physical education. *European Physical Education Review* 21(4), 485-503, DOI: 10.1177/1356336X15584087.
- Dobsen, S. & Engh, R. (red.) (2010). *Vurdering for læring i fag*. Kristiansand: Høyskoleforlaget.

- Dowling, F., Fitzgerald, H. & Flintoff, A. (2012). *Equity and difference in physical education, youth sport and health: A narrative approach*. London: Routledge.
- Dunn, K.E. & Mulvenon, S.W. (2009). A Critical Review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation* 18(7), 1-11.
- Dweck, C.S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.
- Dyson, B. (2014). Quality Physical Education: A Commentary on Effective Physical Education Teaching. *Research Quarterly for Exercise & Sport* 85(2), 144-152.
- Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximise student learning*. Thousand Oaks, CA: Corwin Press.
- Eggen, A., Smith, K. & Dobson, S. (2009). *Vurdering, prinsipper og praksis*. Oslo: Gyldendal akademisk.
- Ellingsen, F., Nilsen, T.A. & Oftedal, T. (1985). *Kroppsøving på ungdomstrinnet. Grunnlag for karaktersetting*. Oslo: Norges idrettshøgskole.
- Elnan, I. & Sando, O.J. (2014). Prestasjonsorientert læringsklima og vurdering i kroppsøvingfaget. I Reinertsen, A.B., Groven, B., Knutas, A. & Holm, A. (red.). *FoU i praksis 2013 conference proceedings* (s. 73-81). Trondheim: Akademika forlag.
- Elstad, E. (2009). Schools which are named, shamed and blamed by media: school accountability in Norway. *Educational Assessment, Evaluation and Accountability* 21(2), 173-189.
- Engelsen, K.S. & Smith, K. (2010). Vurdere egen vurderingspraksis - utvikling av læreres vurderingskompetanse. I Aamotsbakken, B. (red.). *Læring og medvirkning* (s. 75-89). Oslo: Universitetsforlaget.
- Engeström, Y. (1987). *Learning by expanding: an activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.
- Engh, R., Dobsen, S. & Høihilder, E.K. (red.) (2007). *Vurdering for læring*. Oslo: Cappelen Damm.
- Engvik, G. (2010). Elevvurdering i faget kroppsøving – noen utfordringer. I Dobsen, S. & Engh, R. (red.). *Vurdering for læring i fag* (s. 201-213). Kristiansand: Høyskoleforlaget.
- Engvik, G. (2013). Vurderingspraksiser i kroppsøvingfaget. I Sandvik, L.V. & Buland, T. (red.). *Vurdering i skolen. Operasjonaliseringer og praksiser: Delrapport 2 fra prosjektet Forskning på individuell vurdering i skolen (FIVIS)* (s. 83-98). Hentet 21.4.2015 fra: <http://www.udir.no/Upload/Rapporter/2013/FIVIS2.pdf?epslanguage=no>
- Eriksson, C., Gustavsson, K., Quennerstedt, M., Rudsberg, K., Öhman, M. & Öjjen, L. (2005). *Nationella utvärderingen av grundskolan 2003 (NU-03). Idrott och hälsa*. Stockholm: Elanders Gotab.
- Evans, J., Davies, B. & Penney, D. (1999). The Social construction of teaching and learning: The politics of pedagogy. I Hardy, C. & Mawer, M. (red.). *Learning and teaching in physical education*. (s. 9-19). London: Falmer Press.
- Feyerabend, P. (1993). *Against method: outline of an anarchist theory of knowledge (3<sup>rd</sup> edn)*. London: Verso.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4<sup>th</sup> edn)*. London: Sage.
- Folvik, S.E. (1972). *Kroppsøving i grunnskolen. Synspunkter på spørsmål om målsetting og karaktersetting i faget fra 1939 til 1969*. Notater og Rapporter fra Norges idrettshøgskole nr. 4. Oslo: Norges idrettshøgskole.

- Forskrift til opplæringslova. (2006). *Forskrift til opplæringslova*. Hentet fra <https://lovdata.no/dokument/SF/forskrift/2006-06-23-724>
- Gardner, J. (red.) (2012). *Assessment and Learning (2<sup>nd</sup> edn)*. London: Sage.
- Georgakakis, S. & Wilson, R. (2012). Australian physical education and school sport: An exploration into contemporary assessment. *Asian Journal of Exercise & Sport Science*, 9(1), 37-52.
- Gibbons, S.L. & Kankkonen, B. (2011). Assessment as learning in physical education: Making assessment meaningful for secondary school students. *Physical and Health Education*, 6-12.
- Giddens, A. (1976). *New rules of sociological method: A positive critique of interpretative sociologies*. London: Hutchinson.
- Gipps, C. (1994). *Beyond testing: towards a theory of educational assessment*. London: Falmer Press.
- Glaser, R. & Klaus, D.J. (1962). Proficiency measurement: Assessing human performance. I Gagné, R.M. (red.). *Psychological principles in system development* (s. 419-474). New York: Holt, Rinehart & Winston.
- Gravaas, B.C., Hægeland, T., Kirkebøen, L.J. & Steffensen, K. (2008). Skoleresultater 2007 – en kartlegging av karakterer fra grunn- og videregående skoler i Norge. *SSB-notat*. Kongsvinger: Statistisk Sentralbyrå.
- Green, K. (2008). *Understanding physical education*. London: Sage.
- Green, K. (2014). Mission impossible? Reflecting upon the relationship between physical education, youth sport and lifelong participation. *Sport, Education and Society* 19(4), 357–375.
- Græsholt, S.A. (2011). *Elevvurdering og vurderingskultur i kroppsøving*. Masteroppgave. Oslo: Norges idrettshøgskole.
- Grønmo, S. (2004). *Samfunnsvitenskapelige metoder*. Bergen: Fagbokforlaget.
- Gurholt, K.P. & Jenssen, R. (2007). Reformpedagogikkens innpass i kroppsøvingsfaget. *Norsk Pedagogisk Tidsskrift* 91(6), 447-459.
- Hacker, D.J., Dunlosky, J. & Graesser, A.C. (red.) (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.
- Handal, G. & Lauvås, P. (1999). *På egne vilkår: En strategi for veiledning med lærere*. Oslo: Cappelen Akademisk Forlag.
- Hartberg, E.W., Dobson, S. & Gran, L. (2012). *Feedback i skolen*. Oslo: Gyldendal akademisk.
- Harvey, S. & van der Mars, H. (2010). Teaching and assessing racquet games using "Play Practice". Part 1: Designing the right games. *Journal of Physical Education, Recreation and Dance* 81(4), 26-34.
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research* 77(1), 81-112.
- Haug, P. (2003). *Evaluering av Reform 97. Sluttrapport frå styret for Program for evaluering av Reform 97*. Oslo: Norges Forskningsråd.
- Havnes, A., Smith, K., Dysthe, O. & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation* 38(1), 21-27.
- Hay, P.J. (2006). Assessment for learning in physical education. I Kirk, D., Macdonald, D. & O'Sullivan, M. (red.). *The Handbook of Physical Education* (s. 740-751). London: Sage.
- Hay, P. (2010). Systemic inequities in a school-based approach to curriculum and assessment through the eyes of students. *Curriculum Perspectives* 30(1), 14-24.

- Hay, P. & Penney, D. (2009). Proposing conditions for assessment efficacy in physical education. *European Physical Education Review* 15(3), 389–405.
- Hay, P., & Penney, D. (2013). *Assessment in physical education. A sociocultural perspective*. New York, Oxon: Routledge.
- Hay, P. & Macdonald, D. (2008). (Mis)appropriations of criteria and standardsreferenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy and Practice* 15(2), 153–168.
- Hay, P.J. & Macdonald, D. (2010a). Evidence for the social construction of ability in physical education. *Sport, Education and Society* 15(1), 1-18. DOI:10.1080/13573320903217075.
- Hay, P.J. & Macdonald, D. (2010b). The Gendering of Abilities in Senior PE. *Physical Education and Sport Pedagogy* 15(3), 271-285.
- Hay, P., Tinning, R. & Engstrom, C. (2015). Assessment as pedagogy: a consideration of pedagogical work and the preparation of kinesiology professionals. *Physical Education and Sport Pedagogy* 20(1), 31-44. DOI:10.1080/17408989.2013.788145.
- Hensley, L.D., Lambert, L.T., Baumgartner, T.A. & Stillwell, J.L. (1987). Is evaluation worth the effort? *Journal of Physical Education, Recreation and Dance* 58(6), 59-62.
- Hofmeyer, A.T. & Scott, C.M. (2007). Moral geography of focus groups with participants who have preexisting relationships in the workplace. *International Journal of Qualitative Methods*, 6(2), 69-79.
- Hofnes, R. (1993). *Vurdering i kroppsøving. Hva ligger til grunn for læreres karaktersetting i kroppsøving?* Hovedfagsoppgave. Oslo: Norges idrettshøgskole.
- Hofnes, R. (1997). *Elevvurdering i kroppsøving*. Program for skoleforskning. PS-skrift nr. 2/97. Trondheim: NTNU.
- Hofnes, R. (2000). *Elevvurdering i kroppsøving på ungdomstrinnet. Sluttrapport B*. HiST/ALU-Rapport nr. 3. Trondheim: Høgskolen i Sør-Trøndelag.
- Holland, D., Lachicotte, W., Skinner, D. & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.
- Ivanoff, S.D. & Hultberg, J. (2006). Understanding the multiple realities of everyday life: Basic assumptions of focus-group methodology. *Scandinavian Journal of Occupational Therapy* 13(2), 125-132.
- Jacobsen, E.B. (1966). *Kroppsøving som ledd i oppdragelsen i historisk perspektiv*. Hovedfagsoppgave. Oslo: Statens Gymnastikkhøyskole.
- Jacobsen, E.B., Moser, T., By, I.Å., Fjeld, J., Gundersen, K.T. & Stokke, R. (2001). *L97 og kroppsøvingfaget – fra blå praktbok til grå hverdag? Lærernes erfaringer knyttet til den nye læreplanen i kroppsøving*. Hovedrapport 1. Rapport 3/01. Tønsberg: Høgskolen i Vestfold.
- Jacobsen, E.B., Moser, T., By, I.Å., Fjeld, J., Gundersen, K.T. & Stokke, R. (2006). *L97 og kroppsøvingfaget – fra blå praktbok til grå hverdag? Synteserapport*. NFR. Hentet 9.9.2013 fra <http://www.forskningsradet.no/servlet/Satellite?c=Page&pagename=ForskningsradetNorsk%2FHovedsidemal&cid=1175003277667&querystring=jacobsen+kropps%C3%B8ving&spell=true&filters=cssitenam%2C%21ForskningsradetEngelsk%2C%2Clangcodes%2Cno&isglobalsearch=true&configuration=nfrsearchersppublished>
- Jacobsen, A.M. (2010). *Trivsel hos kroppsøvingslærerne i ungdomsskolen: Hvor godt trives kroppsøvingslærerne i ungdomsskolen og hvor sannsynlig er det at de underviser i faget om fem år?* Doktorgradsavhandling. Tromsø: Universitetet i Tromsø.
- James, M. (2006). Assessment, teaching, and theories of learning. I Gardner, J. (red.). *Assessment and learning* (s. 47-60). London: Sage.



- James, M. (2012). Assessment in harmony with our understanding of learning: Problems and possibilities. I Gardner, J. (red.). *Assessment and Learning (2<sup>nd</sup> edn.)* (s. 187-205). London: Sage.
- James, A.R., Griffin, L. & Dodds, P. (2009). Perceptions of middle school assessment: an ecological view. *Physical Education and Sport Pedagogy* 14(3), 323-334.
- Jonskås, K. (2010). *En kunnskapsoversikt over FOU-arbeid innen kroppsøvingfaget i Norge fra januar 1978 – desember 2010*. Oslo: Norges Idrettshøgskole.
- Jonskås, K. (2011). Forskning på kroppsøvingfaget i Norge – hvor står faget? *Kroppsøving* 61(2), 26-28.
- Kamberelis, G. & Dimitriadis, G. (2008). Focus Groups: Strategic articulations of pedagogy, politics and inquiry. I Denzin, N.K. & Lincoln, Y.S. (red.). *Collecting and interpreting qualitative materials* (s. 375-402). Thousand Oaks, CA: Sage.
- Kamberelis, G. & Dimitriadis, G. (2013). *Focus Groups: From structured interviews to collective conversations*. London: Routledge.
- Kirk, D., Macdonald, D. & O'Sullivan, M. (red.) (2006). *The Handbook of Physical Education*. London: Sage.
- Kjerland, G.Ø. (2015). *Å lære å undervise i kroppsøving. Design for utvikling av teoribasert undervisning og kritisk refleksjon i kroppsøvingslærerutdanningen*. Doktorgradsavhandling. Göteborg: Göteborgs universitet.
- Klenowski, V. (2009). Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice* 16(3), 263-268.
- Klenowski, V. & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgments and moderation*. London: Sage.
- Klette, K. (2003). Lærernes klasseromsarbeid; Interaksjons- og arbeidsformer i norske klasserom etter Reform 97. I Klette, K. (red.). *Klasserommets praksisformer etter reform 97* (s. 39-76). Oslo: Pedagogisk forskningsinstitutt.
- Kluger, A.N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119(2), 254-84.
- Folvik, S. E. (1972). *Kroppsøving i grunnskolen. Synspunkter på spørsmål om målsetting og karaktersetting i faget fra 1939 til 1969*. Notater og Rapporter fra Norges idrettshøgskole nr. 4. Oslo: Norges idrettshøgskole.
- KUF (1994). *Læreplan for den videregående skolen, Kroppsøving. Felles allment fag*. Oslo: KUF.
- Kårhus, S. (1992). *Målrelatert evaluering av elever i kroppsøving som felles allment fag*. NIH-Notat. Oslo: Norges idrettshøgskole.
- Lagerstrøm, B. O. (2007). *Kompetanse i grunnskolen. Hovedresultater 2005/2006*. Rapport 2007/21. Oslo/Kongsvinger: Statistisk sentralbyrå.
- Lagerstrøm, B. O., Moafi, H. og Revold, M. K. (2014). *Kompetanseprofil i grunnskolen. Hovedresultater 2013/2014*. Rapport 2014/30. Oslo/Kongsvinger: Statistisk sentralbyrå.
- Larsson, H. & Redelius, K. (red.) (2004). *Mellan nytta och nöje: bilder av ämnet idrott och hälsa*. Stockholm: GIH.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: minute-by-minute and day-by-day. *Educational Leadership*, 63(3), 18-24.

- Leahy, S. & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. I Gardner, J. (red.). *Assessment and Learning (2<sup>nd</sup> edn)* (s. 49-71). London: Sage.
- Leirhaug, P.E. (2008). Høgt karakternivå – del av fagets egenart? *Kroppsøving* 58(2), 16-18.
- Leirhaug, P.E. (2010). Lakmustest for kroppsøvingfaget? *Kroppsøving* 60(5), 32-33.
- Leirhaug, P.E. (2014). (Intet) nytt om kroppsøving i fremtidens skole. *Kroppsøving* 64(5), 16-18.
- Leirhaug, P.E. (2015). Exploring the relationship between student grades and assessment for learning in Norwegian physical education. *European Physical Education Review*, onlinepublikasjon før trykk, DOI:10.1177/136336X15606473.
- Leirhaug, P.E. (2016). Når profesjonaliteten møter seg selv i døra. En historie om innsats, kompetanse og vurdering i kroppsøving. I Grindheim, L.T., Krüger, T., Leirhaug, P.E. & Wilson, D. (red.). *Lærerprofesjonalitet i utdanningspraksiser* (s. 137-156). Bergen: Fagbokforlaget.
- Leirhaug, P.E. & Annerstedt, C. (2015). Assessing with new eyes? Assessment for learning in Norwegian physical education. *Physical Education and Sport Pedagogy*, onlinepublikasjon før trykk, DOI:10.1080/17408989.2015.1095871.
- Leirhaug, P.E. & MacPhail, A. (2015). 'It's the other assessment that is the key': three Norwegian physical education teachers' engagement (or not) with assessment for learning. *Sport, Education and Society* 20(5), 624-640, DOI:10.1080/13573322.2014.975113.
- Leirhaug, P.E., MacPhail, A. & Annerstedt, C. (2016). 'The grade alone provides no learning': investigating assessment literacy among Norwegian physical education teachers. *Asia-Pacific Journal of Health, Sport and Physical Education* 7(1), 21-36, DOI:10.1080/18377122.2016.1145429.
- Leirhaug, P.E. & Midthaugen, P. (2014). Kompetansen hos kroppsøvingslærerne i grunnskolen - fortsatt en utfordring. *Kroppsøving* 64(6), 6-7.
- López-Pastor, V.M., Kirk, D., Lorente-Catalán, E., MacPhail, A. & Macdonald, D. (2013). Alternative assessment in physical education: a review of international literature. *Sport, Education and Society* 18(1), 57-76.
- Lorente-Catalán, E. & Kirk, D. (2016). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review* 22(1), 65-81.
- Lundahl, C. (2011). *Bedömning för lärande*. Stockholm: Norstedts.
- Lyngstad, I. (2013). *Profesjonell kunnskap i skolens kroppsøvingfag: Teoretisk og empirisk belysning i et fenomenologisk og praksisrelatert perspektiv*. Doktorgradsavhandling. Trondheim: NTNU.
- Lyngstad, I., Hagen, P.-M. & Aune, O. (2014). Understanding pupils' hiding techniques in physical education. *Sport, Education and Society*, onlinepublikasjon før trykk, DOI:10.1080/13573322.2014.993960
- Lyngstad, I., Flagestad, L., Leirhaug P.E. & Nelvik, I. (2011). *Kroppsøving i skolen. Rapport fra arbeidsgruppe i kroppsøving* (Udir.-rapport). Hentet 7.6.2012 fra [http://www.udir.no/Upload/larerplaner/forsok/Kroppsoving\\_i\\_skolen\\_rapport\\_060611.pdf](http://www.udir.no/Upload/larerplaner/forsok/Kroppsoving_i_skolen_rapport_060611.pdf)
- Liotard, J.-F. (1988). *The differend: phrases in dispute*. Minneapolis: University of Minnesota Press.
- Lysne, A. (1999). *Karakterer og kompetanse – bind 1: Stridstema i norsk skolehistorie*. Oslo: AVA forlag.

- Lysne, A. (2004). *Karakterer og kompetanse – bind 2: Kampen om skolen 1970-2000*. Oslo: AVA forlag.
- Læreplan for forsøk med 9-årig skole (1960). (1. utgave). Oslo.
- Macdonald, D. (2002). Extending agendas: Physical culture research for the twenty-first Century. I Penney, D. (red.). *Gender and physical education: Contemporary issues and future directions* (s. 208-222). London: Routledge.
- Marsdal, M.E. (2011). *Kunnskapsbløffen*. Oslo: Forlaget Manifest.
- Meld St. nr. 20 (2012-2013). (2013). *På rett vei: Kvalitet og mangfold i fellesskolen*. Hentet fra <https://www.regjeringen.no/contentassets/53bb6e5685704455b06fdd289212d108/no/pdf/s/stm201220130020000dddpdfs.pdf>
- Melograno, V.J. (2007). Grading and report cards for standard-based physical education. *Journal of Physical Education, Recreation and Dance* 78(6), 45-53.
- Mertens, D.M. (1998). *Research methods in education and psychology. Integrating diversity with quantitative and qualitative approaches*. London: Sage.
- Messick, S. (1989). Validity. I Linn, R.L. (red.). *Educational measurement (3<sup>rd</sup> edn)* (s. 13-103). New York: Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23(2), 13-23.
- Messick, S. (1995). Standards of validity and the validity of standards in Performance Assessment. *Educational Measurement: Issues and Practices* 14(4), s. 5-8.
- Miller, M.D., Linn, R.L. & Gronlund, N.E. (2009). *Measurement and assessment in teaching (10<sup>th</sup> Edn)*. Upper Saddle River, NJ: Pearson.
- Mintah, J.K. (2003). Authentic assessment in physical education: prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science* 7(3), 161-174.
- Moe, V.F., Leirhaug, P.E. & Resaland, G.K. (2015). Pedagogisk entreprenørskap i kroppsøvingfagets grenseland. I Haara, F.O. & Ødegård, I.K. (red.). *Grunnskolelærerutdanning gjennom pedagogisk entreprenørskap* (s. 179-200). Oslo: Cappelen Damm Akademisk.
- Moen, K.M. (2011). "Shaking or stirring?" a case-study of physical education teacher education in Norway. Doktorgradsavhandling. Oslo: Norges idrettshøgskole.
- Moen, K.M. & Green, K. (2014). Neither shaking nor stirring: A case-study of reflexivity in Norwegian physical education teacher education. *Sport, Education and Society* 19(4), 415-434.
- Muijs, D. (2011). *Doing quantitative research in education with SPSS (2<sup>nd</sup> edn)*. London: Sage.
- Møller, J. & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Education Administration and History* 45, 336-352.
- Møller, J., Prøitz, T.S. & Aasen, P. (red.) (2009). *Kunnskapsløftet – tung bør å bære. Underveisanalyse av styringsreformen i skjæringspunktet mellom politikk, administrasjon og profesjon*. Rapport 42/09. Oslo: NIFU STEP.
- Nadeau, L., Richard, J.-F. & Godbout, P. (2008). The validity and reliability of a performance assessment procedure in Ice Hockey. *Physical Education and Sport Pedagogy* 13(1), 65-83.

- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist* 22(2), 155-175.
- NESH. (2006). *Forskningsetiske retningslinjer for samfunnsfag, humaniora, juss og teologi*. Oslo: Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora  
Hentet fra <https://www.etikk.no/forskningsetiske-retningslinjer/Samfunnsvitenskap-jus-og-humaniora/>
- Newton, P. (2010). Educational assessment – concepts and issues: the multiple purposes of assessment. I Baker, E., McGaw, B. & Pearson, P. (red.). *International Encyclopedia of Education*. Oxford: Elsevier.
- Ní Chróinín, D. & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: teacher perspectives and experiences. *Physical Education and Sport Pedagogy* 18(2), 219-233.
- NOU 1991:4 (1991). *Veien videre til studie- og yrkeskompetanse for alle*. Oslo: Kirke-, utdannings- og forskningsdepartementet.
- NOU 2002:10. (2002). *Førsteklasses fra første klasse: Forslag til rammeverk for et nasjonalt kvalitetsvurderingssystem av norsk grunnopplæring*. Hentet fra <https://www.regjeringen.no/no/dokumenter/nou-2002-10/id145378/?q=&ch=1>
- NOU 2014:8. (2014). *Elevenes læring i fremtidens skole: Et kunnskapsgrunnlag*. Hentet fra <http://nettsteder.regjeringen.no/fremtidensskole/files/2014/09/NOU201420140007000DDPDFS.pdf>
- NOU 2015:8. (2015). *Fremtidens skole:Fornyelse av fag og kompetanser*. Hentet fra <https://blogg.regjeringen.no/fremtidensskole/files/2015/06/NOU201520150008000DDDPDFS.pdf>
- OECD (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD.
- Ommundsen, Y. (2015). Psykologisk læringsklima i kroppsøving og idrett – betydning for barns og unges læring, trivsel og motivasjon. I Sigmundsson, H. & Ingebrigtsen, J.E. (red.). *Idrettspedagogikk* (s. 46-63). Oslo: Universitetsforlaget.
- Opplæringslova (1998). *Lov om grunnskolen og den vidaregåande opplæringa (opplæringslova)*. Sist endret: LOV-2015-06-19-65 fra 01.10.2015. Hentet fra <https://lovdata.no/dokument/NL/lov/1998-07-17-61>
- Otero-Saborido, F. & González-Jurado, J. (2015). Design and validation of a tool for the formative assessment of invasion games. *Journal of Physical Education and Sport* 15(2), 254-263.
- Ottesen, Y.E. (1994). *Karaktersetting i kroppsøving på ungdomsskoletrinnet. En empirisk undersøkelse blant kroppsøvingslærere*. Hovedfagsoppgave. Levanger: Høgskolen i Levanger.
- Palm, T. (2014). *Assessing with Foucault*. Masteroppgave. Oslo: Norges idrettshøgskole.
- Parker, A. & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research and Method in Education* 29(1), 23-37.
- Peev, O.D. (2001). *Nasjonal standard i vurderingen – realitet, visjon eller luftspeiling? En studie av karaktersettingen i kroppsøving på den videregående skolen i Telemark og Buskerud*. Hovedfagsoppgave. Oslo: Norges idrettshøgskole.
- Penney, D., Brooker, R., Hay, P. & Gillespie, L. (2009). Curriculum, Pedagogy and Assessment: Three message systems of schooling and dimensions of quality physical education. *Sport, Education and Society* 14(4), 421-442.

- Penney, D., Jones, A., Newhouse, P. & Cambell, A. (2012). Developing a digital assessment in senior secondary physical education. *Physical Education and Sport Pedagogy* 17(4), 383-410, DOI:10.1080/17408989.2011.582490.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning. Towards a wider conceptual field. *Assessment in Education: Principles Policy and Practice* 5(1), 85-102.
- Popham, W.J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice* 48(1), 4-11.
- Popham, W.J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator* 46(4), 265-273, DOI:10.1080/08878730.2011.605048.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education* 34(1), 1–20.
- Prøitz, T.S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*, 4(3), DOI:10.3402/edui.v4i3.22629.
- Prøitz, T.S. & Borgen, J.S. (2010). *Rettferdig standpunktvurdering – det (u)muliges kunst? Læreres setting av standpunktkarakter i fem fag i grunnsopplæringen*. Rapport nr. 16/2010. Oslo: NIFU STEP.
- Quennerstedt, M., Annerstedt, C., Barker, D., Karlefors, I., Larsson, H., Redelius, K. & Öhman, M. (2014). What did they learn in school today? A method for exploring aspects of learning in physical education. *European Physical Education Review* 20(2), 282-302.
- Quennerstedt, M., Öhman, M., & Armour, K. (2014). Sport and exercise pedagogy and questions about learning. *Sport, Education and Society* 19, 885–898. DOI:10.1080/13573322.2013.847824.
- Raaen, F.D. (1990). *Elevvurdering i nytt perspektiv. Sluttrapport for prosjektet «Vurdering og veiledning»*. Oslo: Grunnskolerådet.
- Redelius, K. (2007). Betygsättning i idrott och hälsa – en didaktisk utmaning med pedagogiska konsekvenser. I Larsson, H. & Meckbach, J. (red.). *Idrottsdidaktiska utmaningar* (s. 217-231). Stockholm: Liber.
- Redelius, K. & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review* 15(3), 275-294.
- Redelius, K. & Hay, P. (2012). Student views on criterion-referenced assessment and grading in Swedish physical education. *Physical Education & Sport Pedagogy* 17(2), 211-225.
- Redelius, K., Fagrell, B. & Larsson, H. (2009). Symbolic Capital in Physical Education: To Be, to Do or to Know? That is the Gendered Question. *Sport, Education and Society* 14(2), 245–260.
- Redelius, K., Quennerstedt, M. & Öhman, M. (2015). Communicating aims and learning goals in physical education: part of a subject for learning? *Sport, Education and Society* 20(5), 641-655.
- Resch, N. (2009). Justice in grades allocation: teachers' perspective. *Social Psychology of Education* 12(3), 315-325.
- Ringdal, K. (2013). *Enhet og mangfold. Samfunnsvitenskaplig forskning og kvantitativ metode (3. utg)*. Bergen: Fagbokforlaget.
- Risøy, K.E. (2013). *Testing i kroppsøving: Elevenes erfaringer med testing i faget i videregående skole*. Masteroppgave. Oslo: Norges idrettshøgskole.
- Rorty, R. (1989). *Contingency, irony, and solidarity*. New York: Cambridge University Press.
- Ryan, R.M. & Deci, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25(1), 54–67.

- Rønninghaug, M. (2011). *Prestasjonskrav eller treningsglede?: Kroppsøvingfaget i den videregående skole*. Masteroppgave. Bodø: Universitetet i Nordland.
- Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18, 119-144.
- Sadler, D.R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy and Practice* 5(1), 77-85.
- Sandvik, L.V. & Buland, T. (red.) (2014). *Vurdering i skolen. Utvikling av kompetanse og fellesskap: Sluttrapport fra prosjektet Forskning på individuell vurdering i skolen (FIVIS)*. Trondheim: NTNU; SINTEF. Hentet 28.11.2014 fra: <http://www.udir.no/Upload/Forskning/2015/FIVIS%20sluttrapport%20desember%202014.pdf?epslanguage=no>
- Sandvik, L.V., Engvik, G., Fjørtoft, H., Langseth, I.D., Aaslid, B.E., Mordal, S. & Buland, T. (2012). *Vurdering i skolen. Intensjoner og forståelser – Delrapport 1 fra prosjektet Forskning på individuell vurdering i skolen (FIVIS)*. Trondheim: NTNU; SINTEF. Hentet 14.12.2012 fra <http://www.udir.no/Upload/Rapporter/2012/fivis.pdf?epslanguage=no>
- Scriven, M. (1967). The methodology of evaluation. I Tyler, R.W., Gagné, R.M. & Scriven, M. (red.). *Perspectives of curriculum evaluation* (s. 39-83). Chicago, IL: Rand McNally.
- Seger, I. (2014). *Betygsättningsprocess i ämnet idrott och hälsa - En studie om betygsättningsdilemman på högstadiet*. Örebro Studies in Sport Sciences 21, Licentiatuppsats. Örebro: Örebro universitet.
- Skagen, K. (1996). *Karakterboka - Om karakterer og vurdering i ny skole*. Oslo: Universitetsforlaget.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher* 9(7), 4-14. DOI:10.3102/0013189X029007004.
- Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research* 78(1), 153-178.
- Sjøberg, S. (2014). PISA-SYNDROMET: Hvordan norsk skolepolitikk blir styrt av OECD. *Nytt Norsk Tidsskrift* 31(1), 30-43.
- Slemmen, T. (2010). *Vurdering for læring i klasserommet*. Oslo: Gyldendal.
- Skolinspektionen. (2010). *Mycket idrott och lite hälsa*. Rapport 2010-08-30. Hentet 4.4.2015 fra <http://www.skolinspektionen.se/Documents/publikationssok/granskningsrapporter/flygande-inspektioner/2010/idrott/slutrapport-flygande-tillsyn-idrott.pdf>
- Smithson, J. (2008). Focus groups. I Alasuutari, P., Bickman, L. & Brannen, J. (red.). *The Sage handbook of social research methods* (s. 357-430). London: Sage.
- Standal, Ø.F. & Rugseth, G. (red.) (2015). *Inkluderende kroppsøving*. Oslo: Cappelen Damm akademisk.
- Stiggins, R.J. (2002). Assessment Crisis: The Absence of Assessment FOR Learning. *Phi Delta Kappan International* 83(10), 758-765.
- St.meld. nr. 30 (2003-2004). (2004). Kultur for læring. Hentet fra <https://www.regjeringen.no/contentassets/988cdb018ac24eb0a0cf95943e6cdb61/no/pdfs/stm200320040030000dddpdfs.pdf>
- Stobart, G. (2008). *Testing Times. The uses and abuses of assessment*. London/New York: Routledge.
- Stobart, G. (2012). Validity in formative assessment. I Gardner, J. (red.). *Assessment and Learning (2<sup>nd</sup> edn)* (s. 233-242). London: Sage.

- Stokke, K.H., Thronsen, I., Lie, S. & Dale, E.L. (2008). *Evaluering av vurdering for læring: Underveisrapport fra følgeforskningen «Evaluering av modeller for kjennetegn på måloppnåelse i fag»*. Oslo: Universitetet i Oslo, ILS.
- Soltz, S.A. (2014). *The philosophy of physical education. A new perspective*. Oxon, New York: Routledge.
- Svalund, J. (2004). Funksjonshemmede i Europa. *Samfunnsspeilet 1/2004*, 18(1) (s. 34-41). Kongsberg: Statistisk sentralbyrå.
- Svendby, E.B. (2013). *'Jeg kan og jeg vil, men jeg passer visst ikke inn': En narrativ studie om barn og unges erfaringer med kroppsøvingfaget når de har en sjelden diagnose (fysisk funksjonshemming)*. Doktorgradsavhandling. Oslo: Norges idrettshøgskole.
- Svendby, E.B. (2016). (Re)Telling lived experiences in different tales: a potential pathway in working towards an inclusive PE. *Sport, Education and Society* 21(1), 62-81, DOI: 10.1080/13573322.2015.1113166
- Svennberg, L., Meckbach, J. & Redelius, K. (2014). Exploring PE teachers' 'gut-feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review* 20(2), 199-214.
- Synnestvedt, K. (1994). *Skolefaget kroppsøving: fagets bakgrunn og utvikling 1848–1925. En lære- og fagplanhistorisk studie i norsk allmueskole og folkeskole*. Rapport nr. 5. Hovedfagsoppgave. Oslo: Pedagogisk forskningsinstitutt, UiO.
- Säfvenbom, R., Haugen, T. & Bulie, M. (2015). Attitudes toward and motivation for PE. Who collects the benefits of the subject. *Physical Education and Sport Pedagogy* 20(6), 629-646. DOI:10.1080/17408989.2014.892063.
- Tashakkori, A. & Teddlie, C. (red.) (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks: Sage.
- Telhaug, A.O. (2005). *Kunnskapsløftet - Ny eller gammel skole?* Oslo: Cappelen Akademisk Forlag.
- Thompson, M. & Wiliam, D. (2008). Tight but loose: a conceptual framework for scaling up school reforms. I Wylie, E.C. (red.). *Tight but loose: scaling up teacher professional development in diverse contexts* (s. 1-44). Princeton, NJ: Educational Testing Service.
- Thorburn, M. (2007). Achieving conceptual and curriculum coherence in high-stakes school examinations in physical education. *Physical Education and Sport Pedagogy* 12(2), 163-184.
- Thorburn, M. (2008). Articulating a Merleau-Pontian phenomenology of physical education: the quest for active student engagement and authentic assessment in high-stakes examination awards. *European Physical Education Review* 14(2), 263-280.
- Tindall, D. & Enright, E. (2013). Rethinking teacher knowledge in physical education: what do physical education teachers need to know? I Capel, S. & Whitehead, M. (red.). *Debates in physical education* (s. 107-119). London: Routledge.
- Tinning, R. (2012). The Idea of Physical Education: A Memetic Perspective. *Physical Education and Sport Pedagogy* 17(2), 115-126.
- Tolgfors, B. (2014). *Bedömning för lärande i ämnet idrott och hälsa: didaktiska konsekvenser av styrning*. Örebro Studies in Sport Sciences 19, Licentiatuppsats. Örebro: Örebro universitet.
- Tolgfors, B. & Öhman, M. (2015). The implications of assessment for learning in physical education and health. *European Physical Education Review*, onlinepublikasjon før trykk. DOI: 10.1177/1356336X15595006.

- Torrance, H. (2007). Assessment as Learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education* 14(3), 281-294.
- Torrance, H. (2012). Formative assessment at the crossroads: conformance, deformative and transformative assessment. *Oxford Review of Education* 38(3), 323-342. London: Routledge.
- Threlfall, D. (2005). The formative use of assessment information in planning – the notion of contingent planning. *British Journal of Educational Studies* 53(1), 54-65.
- Tveit, S. (2009). Educational assessment in Norway – a time of change. I Wyatt-Smith, C. & Cumming, J. (red.). *Educational Assessment in the 21<sup>st</sup> Century* (s. 227-244). London: Springer.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice* 21(2), 221-237. DOI:10.1080/0969594X.2013.830079.
- Utdanningsdirektoratet (2008). *Undervegsvurdering i fag. Lære meir og betre – kva har vurderingspraksisen til lærarane å seie*. Informasjonsbrosjyre. Oslo: Utdanningsdirektoratet.
- Utdanningsdirektoratet (2011). *Grunnlagsdokument: Satsingen Vurdering for læring 2010-2014*. Oslo: Utdanningsdirektoratet. Hentet 18.6.2013 fra <http://www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Nasjonal%20satsing/2/Grunnlagsdokument%20for%20satsingen%20Vurdering%20for%201%c3%abring%20okt%202011.pdf>
- Utdanningsdirektoratet (2012). *Udir-8-2012 Endringer i faget kroppsøving*. Rundskriv. Hentet 10.9.2012 fra [http://www.udir.no/Regelverk/Rundskriv/2012/Udir82012- Informasjon-om-enderinger-i-fagetkroppsoving-igrunnskolen-og-videregaende-opplaring/](http://www.udir.no/Regelverk/Rundskriv/2012/Udir82012-Informasjon-om-enderinger-i-fagetkroppsoving-igrunnskolen-og-videregaende-opplaring/)
- Utdanningsdirektoratet (2015a). *Endringer i regelverket om vurdering*. Oslo: Utdanningsdirektoratet. Hentet 26.9.2015 fra <http://www.udir.no/Vurdering/Innhold-vurdering/Endringer-i-regelverket-om-vurdering/>
- Utdanningsdirektoratet (2015b). *Revidert læreplan i kroppsøving*. Oslo: Utdanningsdirektoratet. Hentet 21.9.2015 fra <http://www.udir.no/Lareplaner/Finn-lareplan/enderinger/Horing--Justering-av-kompetansomal-i-svømming-i-kroppsøvingsfaget/>
- Utdanningsdirektoratet (u.å.). *Vurdering for læring: Fire prinsipper*. Hentet 8.8.2015 fra: <http://www.udir.no/Vurdering-for-laring/4-prinsipper/>
- van der Mars, H. & Harvey, S. (2010). Teaching and assessing racquet games using "Play Practice". Part 2: Integrating assessment into teaching. *Journal of Physical Education, Recreation and Dance* 81(5), 35-43.
- Vavik, L., Andersland, S., Arnesen, T.E., Arnesen T., Espeland, M., Flatøy., Grønsdal, I., Fadnes, P., Sømoe, K. & Tuset, G.A. (2010). *Skolefagsundersøkelsen 2009. Utdanning, skolefag og teknologi*. HSH rapport 2010/1. Stord: Høgskulen Stord/Haugesund.
- Veal, M.-L. (1988). Pupil assessment practices and perceptions of secondary teachers. *Journal of Teaching in Physical Education* 7(4), 327-342.
- Vingdal, I.M. (red.) (2014). *Fysisk aktiv læring*. Oslo: Gyldendal.
- Vinje, E. (2008a). *Osloundersøkelsen om vurdering i kroppsøving*. Nøtterøy: Ped-media AS.
- Vinje, E. (2008b). Urettferdig vurderingspraksis i kroppsøving. *Kroppsøving* 58(5), 24.
- Vinje, E.E. & Brattenborg, S. (2016). Vurdering i kroppsøvingsfaget – gamle og nye utfordringer. I Vinje, E.E. (red.). *Kroppsøvingsdidaktiske utfordringer* (s. 32-50). Oslo: Cappelen Damm.



- Ward, P. & Lee, M.-A. (2005). Peer-assisted learning in physical education: A review of theory and research. *Journal of Teaching in Physical Education*, 24(3), 205-225.
- Weir, T. & Connor, S. (2009). The use of digital video in physical education. *Technology, Pedagogy and Education* 18(2), 155-171, DOI:10.1080/14759390902992642.
- Whitehead, M. (2010). *Physical literacy throughout the life course*. London: Routledge.
- Wiken, A.R. (2011). *Hva lærer vi i kroppsøving? Hva oppfatter elevene selv at de lærer, og er dette i tråd Kunnskapsløftet 2006?* Masteroppgave. Oslo: Norges idrettshøgskole.
- Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. I Andrade, H.L. & Cizek, G.J. (red.). *Handbook of formative assessment* (s. 18-40). New York, NY: Taylor & Francis.
- Wiliam, D. (2011). What is Assessment for Learning? *Studies in Educational Evaluation* 37, 3-14.
- Wilkinson, S. (2011). Analysing focus group data. I Silverman, D. (red.). *Qualitative Research* (s. 168-184). London: Sage.
- Wilson, D. (2014). *Formativ vurdering gjør forskjell. En empirisk studie av lærernes forståelse av formativ vurdering som vitenskaplig fenomen og politisk konsept*. Doktorgradsavhandling. Trondheim: NTNU.
- Wittgenstein, L. (2003). *Filosofiske undersøkelser*. Oslo: De Norske Bokklubbene.
- Zhu, X. (2015). Student perspectives of grading in physical education. *European Physical Education Review* 21(4), 409-420, DOI:10.1177/1356336X15569628.
- Young, S. (2011). A survey of student assessment practice in physical education. *Strategies: A Journal for Physical and Sport Educators* 24(6), 24-26.

Petter Erik Leirhaug, NIH

## **Artikler og vedlegg**

Oversikt over avhandlingens artikler og vedlegg i den rekkefølge de foreligger.

### **Artikkel 1:**

Assessing with new eyes? Assessment for learning in Norwegian physical education

### **Artikkel 2:**

Exploring the relationship between student grades and assessment for learning in Norwegian physical education

### **Artikkel 3:**

‘The grade alone provides no learning’: Investigating assessment literacy among Norwegian physical education teachers

### **Artikkel 4:**

‘It’s the other assessment that is the key’: Three Norwegian physical education teachers’ engagement (or not) with assessment for learning

### **Vedlegg:**

- 1: Informasjonsskriv med forespørsel til elever om deltaking i spørreundersøkelsen
- 2: Papirversjon av studiens internettbaserte spørreskjema
- 3: Samtykkeerklæring for fokusgruppeintervju
- 4: Temaguide for fokusgruppeintervju
- 5: Godkjenning av prosjektet fra NSD (Norsk samfunnsvitenskapelig datatjeneste – personvernombudet for forskning)
- 6: Innhold i prosjektets databehandleravtale med Websurvey hos NSD (Norsk samfunnsvitenskapelig datatjeneste – personvernombudet for forskning)



**Artikkel 1:**

**Assessing with new eyes? Assessment for learning in Norwegian physical education.**

*Physical Education and Sport Pedagogy* (2015). Elektronisk publisert før trykk 14.10.2015, s.

1–16, DOI: 10.1080/17408989.2015.1095871.



## Assessing with new eyes? Assessment for learning in Norwegian physical education

Petter E. Leirhaug<sup>a\*</sup> and Claes Annerstedt<sup>b</sup>

<sup>a</sup>Norwegian School of Sport Sciences, P.O. Box 4014, Ullevål Stadion, 0806 Oslo, Norway;

<sup>b</sup>Department of Food and Nutrition, and Sport Sciences, University of Gothenburg, Box 300, 405 30 Göteborg, Sweden

(Received 3 January 2015; accepted 18 August 2015)

*Background:* The latest curriculum reform in Norway is one example of an education reform with a highly emphasised assessment for learning (AfL) agenda. Acknowledging that there is a lack of empirical research on AfL in physical education (PE), and that AfL potentially can have an important role to play in development of PE pedagogy, this paper set out to examine the extent to which the emphasis on AfL from educational authorities has led to change in assessment practice in PE.

*Purpose and research question:* The purpose of this paper is to examine the implementation of AfL in PE at upper secondary level in Norway, and discuss possible implications. More specifically we ask ‘How do students’ and teachers’ perspectives of assessment practices in PE reflect AfL key principles?’

*Methods:* A mixed-method design has been applied in this study. Quantitative data, collected through a questionnaire answered by 1486 students from six upper secondary schools (15–19 years), were combined and compared with qualitative data from focus groups of a total of 23 PE teachers at the same schools. Data were analysed in relation to four key principles of AfL.

*Findings:* For the majority of the students in the study, their reports of assessment practice in PE did not reflect the four key principles of AfL. This result was supported by the fact that their PE teachers conveyed very varied understandings and enactments of AfL. The study revealed some difference between teacher and student perspectives regarding AfL key principles, in particular regarding feedback that moves learners forward.

*Conclusions:* The study demonstrates limited implementation of AfL principles in PE and we conclude that the educational authorities’ emphasis of AfL has not proven productive in PE. However, most of the teachers acknowledged the need to change teaching and assessment practices in PE, and all schools in the study are observed to be in an area of changing assessment. Considering the findings on different AfL key principles, this study highlights engaging student more directly in assessment processes as an important development area.

**Keywords:** Formative assessment; AfL; school reform; teacher learning; Norway

### Introduction

A little over 10 years ago, Stiggins deemed the absence of assessment for learning (AfL) in the USA on the political and educational policy level an ‘assessment crisis’, stating that ‘if we are finally to connect assessment to school improvement in meaningful ways, we must

---

\*Corresponding author. E-mail: petter.erik.leirhaug@hisf.no

come to see assessment through new eyes' (Stiggins 2002, 758). Borrowing a distinction that became common in the UK during the 1990s, Stiggins argued that the dominant role of assessment *of* learning, which refers to assessments most often conducted at the end of a teaching and learning process and associated with purposes of accountability, grading, and certification, needed to be balanced with AfL. Assessment research steadily had documented negative implications of some assessment practices, but also showed that, used appropriately as a key part of instruction, assessment carried considerable potential to enhance student achievement (Black and Wiliam 1998; Crooks 1988; Natriello 1987). In a broad sense, AfL is any assessment conducted primarily to support and promote students' learning. More specifically, AfL is identified and understood as classroom assessment practices that share learning goals; involve students in the process and provide them feedback on their own learning; and inform for the purpose of improved learning, teachers' subsequent planning, and pedagogy (Black and Wiliam 1998, 2012; Wiliam 2011). While assessment literature seems to agree on such key principles or characteristics of AfL, there is still ongoing discussion surrounding how to ensure the potential benefits of AfL (Bennett, 2011; Wiliam 2011).

Stiggins' (2002) question of 'assessment crises' already seems a long time ago, as many countries have 'embarked on an education reform with a highly emphasized Assessment for Learning agenda' (Berry 2011, 99). The 2006 Norwegian school reform is no exception. Even if the reform was partly underpinned and justified by accountability perspectives, the new regulations and overall objectives for the assessment were informed by the influential work of Black and Wiliam (1998), and carried an articulated agenda to change the assessment culture and practice in schools in the direction of AfL (Tveit 2014; Utdanningsdirektoratet 2008). Norwegian teachers are supposed to share learning goals and engage students in assessing their own work, and all students are entitled to ongoing assessments ('underveisvurdering' in Norwegian) that promote learning and support their competence development in each subject (Forskrift til opplæringsloven 2006; Tveit 2014).

There has been a remarkable increase in interest in AfL in physical education (PE) literature as well (Hay and Penney 2009, 2013; Hay, Tinning, and Engstrom 2015; López-Pastor et al. 2013; Macdonald 2011; MacPhail and Halbert 2010). The growing interest reflects the constructivist and sociocultural learning theories that developed in the second half of the twentieth century (Hay and Penney 2013) as well as contemporary debates about PE's future as a school subject (Kirk 2010; Tinning 2012). Drawing on Bernstein's (1977) three message systems (i.e. curriculum, pedagogy and assessment), Penney et al. (2009) foreground assessment as one of the fundamental dimensions for understanding and development of quality PE, given that it focuses on intended student learning.

While there is evidence that both the political landscape surrounding education reforms, and more specifically, assessment theory in PE have somewhat embraced AfL, Georgakis and Wilson (2012) believe that an 'assessment crisis' remains in school PE. Interviewing 17 Australian PE and school sport teachers, they highlight a disparity between academic advocacy of AfL and teachers' perspectives. The teachers in their study did not appear to use the term AfL, and the analysis of the interviews did not reveal any AfL-relevant categories. Analysing literature on PE assessment, Georgakis and Wilson (2012) also observed that both research and more practice-orientated textbooks fall short when it comes to providing practical examples of AfL, and argued that their findings may be partially explained by a lack of clarity and consensus on the purposes in PE and school sport. Discussing the validity, reliability, and comparability of PE assessments in Norway, Annerstedt (2010) similarly pointed to unclear subject knowledge in PE as one factor creating problems with the assessments both *of* and *for* learning.



Education and assessment reforms advocating AfL have assumed that implementing AfL into classroom practices would be an unqualified success in terms of enhancing student learning outcomes across different subjects. Although Smith (2011, 55) affirmed that 'AfL is at the heart of the educational discussion in Norway', she concluded that the 2006 reform 'has not succeeded in initiating a major change of common assessment practice in schools' (60). The findings of Sandvik et al. (2012) indicate that this might even be more true for PE compared to other subjects. In comparison to teachers in Norwegian, English, and mathematics, PE teachers report that formative assessment to a lesser extent is integrated into their teaching practice (Sandvik et al. 2012). This suggests that the situation in Norway may not be so different from what Georgakis and Wilson (2012) described in Australia.

### **Purpose and research approach**

Using the 2006 Norwegian reform as an example, the purpose of this paper is to examine the extent to which the emphasis on the implementation of AfL from educational authorities has provided, as Stiggins (2002) poetically put it, new eyes through which to view assessment in PE. Employing data collected from students and PE teachers at six upper secondary schools, the following research question was formulated: How do students' and teachers' perspectives of assessment practices in PE reflect AfL key principles?

In order to answer the research question, we adapted four AfL key principles from Black and Wiliam (1998), previously used by MacPhail and Halbert (2010) to examine AfL in PE: (1) sharing learning intentions with students; (2) sharing criteria for success; (3) involving students in assessing their own (and other students') learning; and (4) providing feedback that moves learners forward.

These principles were formulated to direction teachers in implementing AfL into their practice. In this study, we use them as a framework to analyse both student and teacher data, acknowledging the importance of including students' perspectives. Given the strong student-centered focus that underpins AfL advocacy, it is somewhat ironic that, as an 'agent of change', AfL has mostly focused on the policy and teacher practice levels, while little attention has been paid to students' perspectives (McInerney, Brown, and Liem 2009). Educational authorities and teacher intentions may or may not be available to students, and students can have different perceptions of the same processes. It is only by including students' perspectives and achievements that we are able to make judgements about the validity of AfL (Stobart 2012). Research related to AfL in PE indicates that the implementation of AfL can be productive in terms of teacher learning and can result in positive student learning experiences (López-Pastor et al. 2013; MacPhail and Halbert 2010; Ní Chróinín and Cosgrave 2013). Scaling up from such innovative AfL studies at teacher and classroom level to the implementation of AfL in all schools as part of a nation-wide reform creates new challenges and new questions (Leahy and Wiliam 2012). The 2006 reform in Norway gives us an opportunity to study what happens in PE. Before moving on to more specific methodological considerations, followed by findings, discussion, and conclusion, the next section offers a brief presentation of PE and assessment in Norway in order to contextualise the study.

### **Research context**

The 2006 Norwegian school reform introduced new national curricula with 'competence aims' for all subjects. The competence aims describe the expected learning outcomes in

each subject at different levels and constitute the basis for all student assessment. Although little effort has been made at the national level to inform, explain, and educate the teachers about the new assessment regulations (Tveit 2014), the reform appears to have made an impact on the PE teachers' teaching and assessment practices (Arnesen, Nilsen, and Leirhaug 2013).

PE is a mandatory subject throughout 13 years of school starting at the age of six. The PE curriculum can be described as a multi-activity model in which the students learn about body and health through the use of their bodies. Research, however, indicates that practice in Norwegian PE classrooms is dominated by the teaching of traditional sport skills, and favours students being active in such sports (Säfvenbom, Haugen, and Bulie 2014), not unlike observations in many other countries (Kirk 2010).

The students receive grades twice a year from year eight (age 12) onwards. Even though emphasis in the 2006 reform is placed on formative assessment, enabling the students to get continuous feedback to support their development and learning, the final grades are used as a sorting instrument for entry into higher levels of the education system, and as such, they are high stakes for the students (Tveit 2014). The educational system is highly decentralised in Norway. There are practically no specifications on content and methods in the curriculum, and both the integration of AfL and assigning high-stakes grades to student are, to a large extent, the responsibility of the individual PE teacher.

### **Methods**

The main focus in this study was on AfL implementation in PE at the school level as part of a nation-wide reform. To include both students' and teachers' perspectives, a mixed-method design was applied (Creswell 2009; Tashakkori and Teddlie 2003). Quantitative data, elicited from a questionnaire administered to students in six upper secondary schools in Norway, were compared with results from the analysis of qualitative data from focus groups of the PE teachers in each of the six schools. The methods were chosen to capture a broad representation of assessment in PE at each school, and enable the analysis to discover possible between-school differences. Focus groups using the pre-existing group of PE teachers at each school were appropriate in this study because the main interest was implementation at the school level, and not the practice of the individual teacher. Acknowledging that there can be challenges associated with using pre-existing groups, we also noted that 'the synergy and dynamism generated within homogenous collectives often reveal unarticulated norms and normative assumptions' (Kamberelis and Dimitriadis 2008, 397).

The schools were recruited in a convenient sample (Bryman 2012) with school size, geographic location, and interest in participating in the project as main selection criteria. Seventeen schools were contacted before six had agreed to participate in the research. Data collection was carried out during 2011 and 2012. All participants were assured anonymity and informed of their right to withdraw from the study at any time. The study had informed consent and was ethically approved by the Norwegian Social Sciences Service (NSD).

### **Participants**

The participants in this study were 1486 students (841 girls, 645 boys; response rate 72.6%), and 23 PE teachers (8 females and 15 males) from six upper secondary schools (students aged 15–19 years) in Norway. We have numbered the schools from 1 to 6, and Table 1 provides an overview of the distribution of participants across the schools.

Table 1. Sample of teachers and students in the six schools.

	Number of PE teachers	Population of PE students (N)	Student respondents included in study ( <i>n</i> )	Student response rate (%)
School 1	3	332	283	85.2
School 2	2	169	103	60.9
School 3	7	424	306	72.1
School 4	3	98	87	88.8
School 5	3	254	173	68.1
School 6	6	769	534	69.4
Total	24	2046	1486	72.6

There were a total of 24 teachers teaching PE in the six schools at the time of data collection. Of them, 23 participated in the focus groups; one teacher had fallen sick on the day of the interview. Teachers' ages ranged from 24 to 61 years old. All had at least one year specialist training for PE as part of their teacher education. Three of the teachers held a master's degree in PE.

#### ***Procedures and instrument***

The focus groups were conducted by the first author, and recorded with a digital voice recorder, in a meeting room or classroom at the PE teachers' own school. After asking briefly about the status of PE and the support from school leadership, the researcher presented an open initial question, asking the teachers to describe the assessment practices in PE at their own school. A topic guide based on the PE curriculum, previous research, and AfL theory was used by the researcher to ensure that relevant issues were raised among teachers during the focus groups. Examples of topics were 'understanding of curriculum and assessment regulations', 'PE content', 'assessment for learning', 'grading criteria', and 'the future of PE'.

The questionnaire used to collect quantitative student data was developed to enable researchers to reconstruct a broad picture of students' perspectives of assessment practices at the schools. Informed by the Norwegian PE curriculum, AfL theory, and a set of questions previously used in Sweden (Larsson and Redelius 2004; Redelius and Hay 2012), we developed an online questionnaire to be used in the six schools (examples of items in Tables 2 and 3). In addition to questions about students' perceptions of PE and assessment, the questionnaire had a section on student background and their sport activity outside school. Not being a validated instrument, the online questionnaire was piloted with three different PE classes (67 students). The piloting led to only minor modifications of a few questions. The final questionnaire had 32 questions and took 15–20 minutes for the students to complete.

Students answered the questionnaire in a lesson at their school, most often not a PE lesson. In four of the schools, the first author visited the classroom, presented the study, and emphasised the need for honest responses. At two of the schools, the questionnaire was administered by the regular teacher for each class.

#### ***Data reduction and analyses***

In the student questionnaire, items anticipated to be well suited to inform the research question were identified. In regard to reliability, it was decided that at least two different items

Table 2. Students' responses to statements.

Statement (Results in percent of $n = 1486$ )	1 Totally disagree	2	3	4	5 Totally agree
S1 I know what is needed to perform to receive the different grades	11	18.1	27.3	28.5	15.2
S2 I know what to do to receive a better grade	10.3	14.3	24.5	30.2	20.7
S3 I am being engaged in self-assessment	31.2	25.3	27.1	11.7	4.8
S4 The students are engaged in peer assessment	46.1	25.0	18.6	7.1	3.2
S5 I receive feedback on where I am in the learning process.	25.0	27.7	30.5	12.5	4.3

Table 3. Students' responses to questions.

Question (Results in percent of $n = 1486$ )	Never	1 time	2–4 times	5–10 times	More than 10 times
Q1 How many times this year has the teacher given you feedback on how to improve your competence in PE?	31.6	36.5	27.1	3.3	1.4
Q2 How many times, this school year, have you (in class) spoken about the competence aims in PE?	16.5	39.0	35.3	6.8	2.4
Q3 How many times, this school year, have you assessed your own work in PE?	51.1	26.7	18.5	3.0	0.7

would be obtained to inform each of the four AfL key principles. In total, nine items were elicited. Five of these were statements, in which students had responded to a five-point Likert scale from 'totally disagree' to 'totally agree' (cf. Table 2). The other four items were questions, of which three are presented in Table 3. The fourth question was 'Do you know the competence aims in PE?'

The items included in the study were screened for missing, outliers, and implausible values, and found to be normally distributed. The descriptive statistics of each item were examined and the items are presented as percent frequencies in the findings section. To detect between-school differences, one-way ANOVA tests were conducted for all individual items. Due to differences in sample size and possible violation to assumption of homogeneity of variance, Welch's  $F$ -ratio and Brown–Forsythe  $F$ -ratio were also tested. Eta squared was calculated as a measure of effect size and was interpreted in the light of the classification suggested by Cohen (1988). In *post hoc* tests, Games–Howell procedure was used to identify which, if any, schools differed significantly ( $p < .05$ ).

Focus groups lasted from 45 to 101 minutes. Total recorded time was 7 h 2 min. The focus groups were transcribed verbatim and, following our primary research interest, subjected to a theoretical guided thematic analysis (Braun and Clarke 2006; Bryman 2012). First we read and reread the focus group transcripts in order to code and separate information according to the four AfL key principles. Then, the selected pieces of text were compared and interpreted to identify similarities and differences in knowledge, understanding, and implementation of AfL key principles across schools as well as between teachers. Finally, the results from analysis of teacher data on each AfL principles were compared with findings in student data in order to examine the extent to which students' and teachers'

perspectives were congruent. We also utilised between-school differences revealed in the analysis of student data (i.e. one-way ANOVA) to direct a systematic rereading of teacher data in search for parallel differences.

Before moving to the discussion, the next section reports the findings structured according to the analytical procedures and the AfL key principles. The findings are presented in this way to highlight potential differences between students' and teachers' perspectives regarding each of the four AfL principles, and to ease discussion of how the perspectives reflect the principles.

### Findings

Findings from teacher data are presented in relation to the results from student data and sorted under subheadings following the four AfL key principles. Before that, Tables 2 and 3 provide an overview of included items from the questionnaire and the frequency (in percent) of students' responses. In the left column each item is provided with a number, which will be used as reference in further presentation. The subsection for each AfL principles starts with comments on results reported in Tables 2 and 3, and then subsequently add findings from the analysis of teacher data from focus groups.

Using a one-way ANOVA to analyse between-school differences revealed statistically significant differences between some schools for all items in Tables 2 and 3 (the statistical results in the Appendix). *Post hoc* comparisons indicated that none of the six schools were significantly different when analysis included the four AfL key principles all together. Interpretation of *post hoc* comparisons of each individual variable and consideration of the effect sizes (ranged from  $\eta^2 = .03$  to  $\eta^2 = .07$ , see the Appendix), also taking visual analysis of mean plots into account, made us focus on three main between-schools differences: students in school 2 reported discussing the competence aims more often than students in the other five schools. Schools 2 and 6 appear to have considerably lower mean scores in regard to knowledge of grades and grading criteria compared to the other schools, and this also was the case for schools 3 and 6 in regard to feedback. These between-school differences from student data were also used to consider parallel between-school differences in teachers' descriptions and reflections on assessment practice. The result of this analysis is presented under the AfL key principle subheadings.

#### *Sharing learning intentions with students*

On the question, 'Do you know the competence aims in PE?' 526 students (35.4%) responded, 'No, I do not know them'. Another 207 students (13.9%) reported only having read them on their own. In other words, half of the students in the study reported that the teachers do not share the learning intentions of the curriculum with them. This is supported by question Q2 in Table 2, where the majority of the students reported that they never or only once this year have spoken about the competence aims in PE. 'Only once' probably is best interpreted as the first lesson of the year, when the students usually get general information about the subject and how it is planned to be organised throughout the school year.

The teachers in all schools demonstrated knowledge about the competence aims in the national curriculum and awareness of changes in assessment regulations. How they reflected upon and described communication with students about the competence aims were completely another issue. Most teachers seemed to think that the competence aims

were too abstract or vague to be useful to provide feedback, and that the students are not able to grasp the expected learning outcomes directly from the competence aims. In school 6, one teacher clearly enough explained, ‘The competence aims are a super structure. You certainly have to set up more concrete learning goals under them’.

In two schools the teachers themselves referred directly to AfL when explaining how they worked with the competence aims. School 2, in which the analysis of one-way ANOVA and *post hoc* tests found students to more often discuss the competence aims, was one of these. The discussion between the two PE teachers in school 2 confirms that they, compared to the teachers in the other schools, tried to make the role of the competence aims more visible to students. One of the teachers said:

What’s important is that the students know what is expected of them—that they know what they are going to be measured in relation to. So in advance, at the beginning of a teaching period, we present the competence aims from the syllabus, and so, at least I, try to break down the competence aim into learning goals, then, it all becomes a little more concrete for students.

In school 2, the teachers also referred to many instances in which students disagreed or discussed assessment results, and explained how they then referred to the regulations and competence aims:

Often they [students] are unhappy and they come and complain and do not understand why they got grade 3 or 4, even if, as my colleague said, they did know the goals that they were expected to achieve.

### ***Sharing criteria for success***

The students’ distribution in regard to questions S1 and S2 in Table 2 illustrates that when it comes to grading, about half of the students report being familiar with the success criteria. Nevertheless, about a quarter of the students report the opposite – that they do not know the grading criteria and what to do to receive a better grade in PE.

PE teachers in all six schools expressed awareness of the importance of sharing the success criteria with students. Some teachers expressed in a direct manner that they tried ‘to be very clear on this’. Most of the changes in assessment practice, which the PE teachers described, were connected to making the criteria for assessment, and grading in particular, more visible to the students:

The idea was that they [students] should know that there would be some predictability. That they should be familiar with the assessment criteria and assessment situations.

Analysis of one-way ANOVA and *post hoc* test revealed students’ reports in schools 2 and 6 to be different from the others. We have tried to trace this in the focus groups without success. There is no obvious difference in the way teachers in these two schools described and discussed the importance of clarifying criteria and how to enact this, compared to discussions among teachers in the other schools. Noteworthy, however, is that in the other four schools the teachers made explicit references to the fact that the students seemed satisfied; they did not receive many discussions or complaints about assessment and grades. This indicates that there might be differences not revealed in our data regarding how success criteria are framed and in fact presented and enacted by the teachers.

***Involving students in assessing their own (and other students') learning***

On question Q3, 51.1% of the students reported to never have assessed their own work in PE. Another quarter reported to have done this only once. Comments in focus groups demonstrate self-assessment as something teachers were exploring as a way of developing their own assessment practice. Nevertheless, self-assessment seemed in most cases attached to students assigning grades to own performances, subsequently used as a means to 'regulate reality orientation', as a teacher in school 6 put it, of the student's awareness of his or her level. Further, the examples provided by the teachers were often connected to a competence aim in the PE curriculum stating that a student should be able to 'plan, carry out and assess a personal training period'.

The reports from students on statement S3 confirm the picture of self-assessment as not a very regular practice in PE, and statement S4 shows that about 70% of students report that they do not experience the teachers engaging students in peer assessment. This coincides with results from teacher data. While the PE teachers seemed to value, and some of them explored self-assessment practices in their own teaching, they were reluctant towards peer assessment. Prompted with a question about peer assessment, teachers in only one school have engaged with this. A straightforward answer in another school in fact was quite illustrative: 'No, I have never done that'. Several teachers in other schools expressed their scepticism about facilitating peer assessments, reflecting that it might be difficult for students to be honest in regard to their peers.

***Providing feedback that moves learners forward***

Responses to question Q1 and statement S5 show that most students do not experience feedback so often that it becomes meaningful in a learning perspective. Having received feedback from the teacher on how to improve their competence in PE only once or not at all, which over 60% of the students reported, does not qualify as ongoing feedback regarding the next step in the learning process.

In stark contrast to this, the teachers conveyed the view that they 'give feedback all the time' during PE lessons. A good illustration of what they meant by this is the following explanation:

If you could log all the comments given in the lessons throughout a year, I think they all could become meaningful in the light of some of the competence aims. But, perhaps, I'm not good enough at communicating that all these feedback are part of the on-going assessment in PE, and feedback in relation to the overall competence aims.

Some teachers quite clearly articulated the inherent challenge in communicating this to students:

The students expect that, just as in other subjects, to sit down, with a paper, talking. You have no chance to do this as a PE teacher. . . . They do not realize that when I say 'watch your head', 'lift your head', 'drop your shoulders', [they are] constantly getting feedback.

The teachers presented this way of thinking about feedback as somewhat subject-specific, and that there probably was a lack of understanding of this by those who are not involved in teaching PE: 'I think this is something we are not good enough to communicate out of our professional group'. The main challenge seemed not related to provide good and

meaningful feedback, but rather how to make the already existing practice of feedback visible to students (and others not engaged in teaching PE).

The one-way ANOVA and *post hoc* test showed students in schools 3 and 6 reporting significantly less feedback compared to the other schools. We were not able to trace the between-schools difference revealed in student data when comparing teachers' discussions in the focus groups. We do, though, note that schools 3 and 6 were the two largest schools in our study.

## Discussion

The four AfL key principles were identified in both students' and teachers' perspectives of assessment practices in PE across all schools. When all four key principles were simultaneously included in the analysis of between-school differences, none of the six schools was significantly different from the others. This may signify that the 2006 Norwegian reform's emphasis on AfL has led to somewhat similar impacts at the single-school level.

However, the majority of students reported very few or no experiences regarding three of the AfL key principles, and it was only in regard to the fourth – sharing criteria for success – that a larger number of students reported high scores. Therefore, the students' perspectives of assessment practices in PE cannot, in general, be considered to reflect the AfL principles. The analysis of teacher data revealed that most PE teachers had only recently begun to explore formative assessment practices, and most possessed little experience, and some none, in integrating AfL into their teaching. This finding coincides with and partly explains the students' low reports of AfL experiences.

Demonstrating that half of the students are precluded from the potential benefits of AfL in PE, our findings indicate that official encouragement from educational authorities to engage in AfL has not yet proven productive for all students. On the contrary, it can be argued that efforts to implement AfL have showed a potential to marginalise a relatively large group of students and increase differences between low and high achievers in PE. Without help to understand the purposes of classroom assessment and explicit instruction on how to use feedback and self-assessment in their learning processes in PE, the students who benefit from such AfL strategies may be the exception rather than the rule. Therefore, it becomes the responsibility of teachers to create assessment tasks and dialogues that include all students and help them understand the criteria. The PE teachers in this study did not appear to successfully accomplish this, even if, in their own views, they emphasised clarity in sharing assessment criteria. While this effort seemed to reach the majority of students, another quarter of the students still disagreed that this was clear.

School 2 is particularly interesting considering that students in this school reported more knowledge of the competence aims and less about the success criteria as compared to the other schools. With the understanding that the items used in sharing of success criteria concentrated on grading, School 2's results may indicate that the assessment and shared grading criteria are not well aligned with the competence aims and assessment regulations of the national curriculum. Therefore, this reiterates a well-known challenge with assessment validity in PE (Hay and Macdonald 2008; Redelius and Hay 2012). However, in contrast to extensive research that has demonstrated validity problems due to misalignment of assessment, pedagogy, and curriculum (Georgakis and Wilson 2012; Hay and Penney 2013), as well as PE teachers using their 'gut feeling' or internal criteria when assessing and grading students (Hay and Macdonald 2008; Svennberg, Meckbach, and Redelius 2014), the teachers in all schools in this study referred to the national curriculum and the competence aims as a basis for assessment and criteria.



This fosters a more positive interpretation of students' low reports on AfL. Although the PE teachers expressed varied and mostly superficial understandings of AfL, an important finding from this study is that teachers seemed to recognise changing assessment practices as a way of developing their teaching practice, as well as how it can contribute to making PE a subject "fit" for the contemporary environment' (Tinning 2012, 123). However, they do not yet realise how fundamental this may be to their pedagogical thinking and practice. Black and Wiliam (2012) hinted that formative assessment and AfL are more to be regarded as theories of pedagogy rather than of assessment. Hay et al. (2015) support this view and have recently advocated its significance in the context of PE. The concepts of AfL probably require most PE teachers to teach in ways they have not taught before, and in ways their students have not experienced. Embracing AfL will then require unlearning some of the practices and beliefs that have dominated teachers' careers to date as well as helping students to take on new roles in their learning.

Considering the lack of practical or instructional examples of how to use assessment to improve learning in PE, noted by Georgakis and Wilson (2012), we will highlight the need for PE subject-specific development of pedagogical tools and strategies regarding AfL. Development and exploration of AfL tools and strategies designed for PE should be regarded as professional learning activities. As indicated in earlier research (MacPhail and Halbert 2010; Ní hróinín and Cosgrave 2013), at its best, implementation of AfL enhances the focus of learning and feeds the critical awareness of what moving activities and learning outcomes are valuable in PE practice. This study suggests that such positive benefits do not necessarily follow when scaling up from AfL implementation at teacher and classroom level to the implementation of AfL into all classrooms as part of a nationwide assessment reform. Most PE teachers in the study had not grasped the deeper understanding of AfL. The study showed few opportunities to share and moderate assessment reflections, and nothing like 'a professional development program specifically designed to give teachers the expertise they need to assess for learning' (Stiggins 2002, 765), exist in the Norwegian PE teachers' environment. Without a framework of ongoing professional development, there is a risk that, in the end, AfL instructional tools and strategies, as observed in studies by Gardner et al. (2011, 111), 'soon became drab routine for some teachers and no longer met the ever-changing needs of the classroom'.

The findings revealed a degree of incongruence between teachers' and students' reports on sharing learning intentions, but the importance of including both perspectives is most obvious in relation to reports on feedback. The teachers emphasised that they gave feedback continuously in PE, while the majority of students report not receiving feedback. This is an invitation to rethink Black and Wiliam's (2012) anticipation of feedback being easier to bring forward in PE and other such more clear performance subjects, 'in that its purpose can be evident to both teacher and student, and it is clear that the learning is entirely dependent on it' (Black and Wiliam 2012, 223). What was evident for the PE teachers regarding feedback in our study was not evident for the students. However, this contradiction is understandable when we consider that the teachers indirectly expressed being satisfied with their feedback, but not with their communication of how all small tips and comments given during practice are in fact feedback according to the competence aims. The result also revealed students in the largest schools, number 3 and 6, reporting less feedback than students in the other schools. Not finding similar differences in teacher data, this can indicate that communication, if that is the problem, is even more difficult at larger schools. Central to the argument from PE teachers about feedback is the question: can feedback move learning forward without being visible and understood by students as feedback? It probably can, but may not contribute to assessment as pedagogy (Hay et al. 2015) or to creating

self-regulated learners who leave school confident to continue movement learning and activity throughout their lives.

Implementation of the AfL principle related to self- and peer-assessment is still largely absent in both students' and teachers' reports on assessment practice. While speaking positively about self-assessment, only some of the teachers actually have used it with their students. Peer assessment was even more rare, and appeared to be one area the PE teachers were sceptical about. This finding contradicts to those of Topping (2013), who noted teachers' haste to involve students in peer assessment before preparing them to complete such tasks productively. These more negative views may be associated with the body and some subject-relevant performances being more visible – perhaps more personal, even – in PE compared to other school subjects. However, students reconstruct subject-relevant learning and assessment criteria in ways that can differ extensively from intentions of teachers and in curriculum (Redelius and Hay 2012), and greater student involvement through AfL implementation, particularly self- and peer-assessment, is perhaps one way to face these challenges. Investigating the implications of AfL, Tolgfors and Öhman (2015) showed that, if teaching practices in PE are open to student influence and resistance, there will be a constant negotiation between students and teacher about the subject content.

Another concern emerging from the teachers' explanation of feedback in PE understood as comments and tips on optimising movement is how this easily fosters a reflection of what Kirk has named 'physical education-as-sport-techniques' (2010, 4). Here, practices are directed towards learning and performing concrete skills from known sports, in which some ways of moving are predetermined to be more correct than others. This gives students little opportunity to own their own learning, and assessment becomes restricted to a deficit model of what students cannot (yet) do according to the predetermined sport and skill expectations. In contrast, MacPhail and Halbert (2010) stress the importance of rich tasks in PE to build an AfL-friendly environment. Being confident that they are good at providing feedback in PE practice, and taking more or less for granted that this feedback aligns with the broad specter of competences and learning outcomes in the national curriculum, the teachers risk not grasping what can be both a serious validity problem and a hindrance for deeper understanding of AfL.

### **Conclusion and direction for development**

Using the Norwegian 2006 reform as one example of an educational reform with an emphasised AfL agenda, this study's main purpose was to examine the implementation of AfL in PE. Student and teacher data from six upper secondary schools in Norway were analysed in relation to four key principles of AfL adapted from the literature (Black and Wiliam 1998; MacPhail and Halbert 2010).

The findings coincide with earlier observations of assessment in PE (Georgakis and Wilson 2012; López-Pastor et al. 2013) and demonstrates limited implementation of AfL principles in PE. The majority of the students reported not being engaged in assessment of their own and others' work in PE, and never or only once having received feedback from the teachers about progress regarding the competence aims in the national PE curriculum. The students' reported lack of experience with AfL principles was supported by the fact that their PE teachers conveyed very varied understandings and enactments of AfL principles, and of the roles and purposes of assessment in general. Although most teachers in the study explored new strategies for student assessment – partly motivated and pressured by the 2006 reform assessment and documentation requirements – those strategies were not clearly embedded in learning theory or applied with an articulated focus

on improving teaching and learning in PE. Therefore, we conclude that the PE teachers did not possess the essential understanding of, or competence in, AfL to successfully integrate it into their teaching.

If the 2006 reform's assessment agenda, and what Bennett (2011, 21) in his critical review called the 'considerable promise' of AfL, are to transition from a 'talked about' theory of innovations to a sustainable 'actioned' pedagogical future in PE (Casey 2013, 79), this study indicates a need to raise assessment literacy among Norwegian PE teachers and provide practical examples of how to translate AfL concepts into PE classrooms. Research on the implementation of AfL has shown it to be a time-consuming process (Gardner et al. 2011; Klinger, Volante, and DeLuca 2012), and PE teachers will probably need considerable time, and some sort of professional development support, to explore and evaluate the benefits of different AfL strategies in their own teaching. In addition, research is required to ensure the validity and fairness of these assessments, as well as to track positive and avoid negative impacts. As noted by Tolgfors and Öhman (2015), implementation of AfL key principles can be reconstructed to support both control and freedom in the learning process in PE, and lead to quite different implications for students.

One particular contribution of this study is a clear call to involve students more directly in the assessment processes. Assessing with new eyes represents a move away from assessment as something done 'to students' to assessment as a pedagogical tool teachers can use 'with students'. It may be that the students' voices about their experiences in PE classrooms create a space where teachers and students can participate in a dialogue about what is relevant and worthwhile learning in PE. Moreover, they could together explore how to elicit evidence of such learning, as well as how to use that evidence to make decisions about the next step in the instruction and learning processes. Such dialogues are not only a prerequisite for successful implementation of AfL, but will likely motivate students and require more clarity from PE teachers (and their students) about learning intentions, preventing them from reliance on 'gut feeling' in student assessment. In this way, AfL may have potential to help transform PE into a subject that provides meaningful and transferable learning experiences for all students, which today it often fail to do (Kirk 2010; Säfvenbom et al. 2014; Williams, Hay, and Macdonald 2012).

### Acknowledgements

We thank two anonymous reviewers for critically reading an earlier version of the manuscript and suggesting substantial improvements as well as providing most helpful detailed comments.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### References

- Annerstedt, C. 2010. "Karaktersetting i kroppsøving: Problematisk, urettferdig og neppe likeverdig [Grading in PE: Problematic, Unfair and Probably Inequal]." In *Aktive liv: Idrettspedagogiske perspektiver på kropp, bevegelse og dannelse*, edited by K. Steinsholt and K. P. Gurholt, 233–254. Trondheim: Tapir.
- Arnesen, T. E., A. K. Nilsen, and P. E. Leirhaug. 2013. "'Den læreplanen som ikkje kan tilpassast mi undervisning, finst ikkje.' Vurdering og undervisning i kroppsøving etter Kunnskapsløftet [Assessment and Teaching in Physical Education after the Knowledge Promotion Act]." *Tidsskriftet FOU i praksis* 7 (3): 9–32.

- Bennett, R. E. 2011. "Formative Assessment: A Critical Review." *Assessment in Education: Principles, Policy and Practice* 18 (1): 5–25.
- Bernstein, B. 1977. *Class Codes and Control: Towards a Theory of Educational Transmissions*. Vol. 3. London: Routledge and Keegan Paul.
- Berry, R. 2011. "Assessment Reforms Around the World." In *Assessment Reform in Education: Policy and Practice*, edited by R. Berry and B. Adamson, 89–102. Dordrecht: Springer.
- Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy, and Practice* 5 (1): 7–74.
- Black, P., and D. Wiliam. 2012. "Developing a Theory of Formative Assessment." In *Assessment and Learning*, 2nd ed., edited by J. Gardner, 206–229. London: Sage.
- Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101.
- Bryman, A. 2012. *Social Research Methods*. 4th ed. Oxford: Oxford University Press.
- Casey, A. 2013. "Practitioner Research: A Means of Coping with the Systemic Demands for Continual Professional Development?" *European Physical Education Review* 19 (1): 76–90.
- Cohen, J. W. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Creswell, J. W. 2009. *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. London: Sage.
- Crooks, T. J. 1988. "The Impact of Classroom Evaluation Practices on Students." *Review of Educational Research* 58 (4): 438–481.
- Forskrift til opplæringsloven. 2006. Forskrift til opplæringsloven [last change: FOR-2014–07–01–987]. Accessed April 28, 2015. <https://lovdata.no/dokument/SF/forskrift/2006-06-23-724?q=forskrift+oppl%C3%A6ringslova>.
- Gardner, J., W. Harlen, L. Hayward, and G. Stobart. 2011. "Engaging and Empowering Teachers in Innovative Assessment Practice." In *Assessment Reform in Education: Policy and Practice*, edited by R. Berry and B. Adamson, 105–119. Dordrecht: Springer.
- Georgakis, S., and R. Wilson. 2012. "Australian PE and School Sport: An Exploration into Contemporary Assessment." *Asian Journal of Exercise and Sport Science* 9 (1): 37–52.
- Hay, P., and D. Macdonald. 2008. "(Mis)appropriations of Criteria and Standards-referenced Assessment in a Performance-based Subject." *Assessment in Education: Principles, Policy and Practice* 15 (2): 153–168.
- Hay, P., and D. Penney. 2009. "Proposing Conditions for Assessment Efficacy in PE." *European Physical Education Review* 15 (3): 389–405.
- Hay, P., and D. Penney. 2013. *Assessment in Physical Education. A Sociocultural Perspective*. New York: Routledge.
- Hay, P., R. Tinning, and C. Engstrom. 2015. "Assessment as Pedagogy: A Consideration of Pedagogical Work and the Preparation of Kinesiology Professionals." *Physical Education and Sport Pedagogy* 20 (1): 31–44. doi:10.1080/17408989.2013.788145.
- Kamberelis, G., and G. Dimitriadis. 2008. "Focus Groups: Strategic Articulations of Pedagogy, Politics and Inquiry." In *Collecting and Interpreting Qualitative Materials*, edited by N. K. Denzin and Y. S. Lincoln, 375–402. Thousand Oaks, CA: Sage.
- Kirk, D. 2010. *Physical Education Futures*. London: Routledge.
- Klinger, D. A., L. Volante, and C. DeLuca. 2012. "Building Teacher Capacity Within the Evolving Assessment Culture in Canadian Education." *Policy Futures in Education* 10 (4): 447–460.
- Larsson, H., and K. Redelius, eds. 2004. *Mellan nytta och nöje: bilder av ämnet idrott och hälsa* [Between Benefit and Joy: Pictures of the Subject PE and Health]. Stockholm: GIH.
- Leahy, S., and D. Wiliam. 2012. "From Teachers to Schools: Scaling up Professional Development for Formative Assessment." In *Assessment and Learning*, 2nd ed., edited by J. Gardner, 49–71. London: Sage.
- López-Pastor, V. M., D. Kirk, E. Lorente-Catalán, A. MacPhail, and D. Macdonald. 2013. "Alternative Assessment in PE: A Review of International Literature." *Sport, Education and Society* 18 (1): 57–76.
- Macdonald, D. 2011. "Like a Fish in Water: Physical Education Policy and Practice in the Era of Neoliberal Globalization." *Quest*, 63: 36–45.
- MacPhail, A., and J. Halbert. 2010. "We Had to Do Intelligent Thinking During Recent PE: Students' and Teachers' Experiences of Assessment for Learning in Post-primary PE." *Assessment in Education: Principles, Policy and Practice* 17 (1): 23–39.

- McInerney, D. M., G. T. L. Brown, and G. A. D. Liem, eds. 2009. *Student Perspectives on Assessment: What Students Can Tell Us About Assessment for Learning*. Charlotte, NC: Information Age Publishing.
- Natriello, G. 1987. "The Impact of Evaluation Processes on Students." *Educational Psychologist* 22 (2): 155–175.
- Ní Chróinín, D., and C. Cosgrave. 2013. "Implementing Formative Assessment in Primary Physical Education: Teacher Perspectives and Experiences." *Physical Education and Sport Pedagogy* 18 (2): 219–233.
- Penney, D., R. Brooker, P. Hay, and L. Gillespie. 2009. "Curriculum, Pedagogy and Assessment: Three Message Systems of Schooling and Dimensions of Quality Physical Education." *Sport, Education and Society* 14 (4): 421–442.
- Redelius, K., and P. Hay. 2012. "Student Views on Criterion-referenced Assessment and Grading in Swedish PE." *Physical Education and Sport Pedagogy* 17 (2): 211–225.
- Säfvenbom, R., T. Haugen, and M. Bulie. 2014. "Attitudes Toward and Motivation for PE. Who Collects the Benefits of the Subject." *Physical Education and Sport Pedagogy*. doi:10.1080/17408989.2014.892063.
- Sandvik, L. V., G. Engvik, H. Fjørtoft, I. D. Langseth, B. E. Aaslid, S. Mordal, and T. Buland. 2012. *Vurdering i skolen. Intensjoner og forståelser* [Assessment in School. Intentions and Understandings]. Research report. Trondheim: NTNU, SINTEF. Accessed December 14, 2012. <http://www.udir.no/Upload/Rapporter/2012/fivis.pdf?epslanguage=no>.
- Smith, K. 2011. "Professional Development of Teachers – A Prerequisite for AfL to Be Successfully Implemented in the Classroom." *Studies in Educational Evaluation* 37: 55–61.
- Stiggins, R. J. 2002. "Assessment Crisis: The Absence of Assessment FOR Learning." *Phi Delta Kappan International* 83: 758–765.
- Stobart, G. 2012. "Validity in Formative Assessment." In *Assessment and Learning*, 2nd ed., edited by J. Gardner, 233–242. London: Sage.
- Svennberg, L., J. Meckbach, and K. Redelius. 2014. "Exploring PE Teachers' 'Gut-feelings': An Attempt to Verbalise and Discuss Teachers' Internalised Grading Criteria." *European Physical Education Review* 20: 199–214.
- Tashakkori, A., and C. Teddlie, eds. 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks: Sage.
- Tinning, R. 2012. "The Idea of Physical Education: A Memetic Perspective." *Physical Education and Sport Pedagogy* 17 (2): 115–126.
- Tolgfors, B., and M. Öhman. 2015. "The Implications of Assessment for Learning in Physical Education and Health." *European Physical Education Review*. doi:10.1177/1356336X15595006.
- Topping, K. J. 2013. "Peers as a Source of Formative and Summative Assessment." In *SAGE Handbook of Classroom Assessment*, edited by I. J. McMillan, 395–412. Los Angeles, CA: Sage.
- Tveit, S. 2014. "Educational Assessment in Norway." *Assessment in Education: Principles, Policy & Practice* 21 (2): 221–237. doi:10.1080/0969594X.2013.830079.
- Utdanningsdirektoratet. 2008. *Undervegsvurdering i fag. Lære meir og betre – kva har vurderingspraksisen til lærarane å seie* [Formative Assessment. Learning More and Better – What Is the Impact of Teachers' Assessment Practice]. Information document. Oslo: Utdanningsdirektoratet.
- William, D. 2011. "What is Assessment for Learning?" *Studies in Educational Evaluation* 37: 3–14.
- Williams, B. J., P. J. Hay, and D. Macdonald. 2012. "The Outsourcing of Health, Sport and Physical Educational Work: A State of Play." *Physical Education and Sport Pedagogy* 16 (4): 399–415.

## Appendix

The between-schools one-way ANOVA revealed statistically significant differences according to the  $p < .05$  level for all individual variables. The items and coding used are presented in Tables 2 and 3 in the paper. Levene's test indicated inhomogeneity of variances, therefore we chose to report Welch's  $F$ -ratio with adjusted degrees of freedom for the residuals for the dependent variables: S1:  $F(5, 410) = 10.48, p < .001, \eta^2 = .03$ ; S2:  $F(5, 413) = 12.07, p < .001, \eta^2 = .04$ ; S3:  $F(5, 410) = 9.91, p < .001, \eta^2 = .03$ ; S4:  $F(5, 406) = 14.06, p < .001, \eta^2 = .05$ ; S5:  $F(5, 404) = 17.55, p < .001, \eta^2 = .05$ ; Q1:  $F(5, 406) = 22.97, p < .001, \eta^2 = .07$ ; Q2:  $F(5, 405) = 17.49,$

$p < .001$ ,  $\eta^2 = .07$ ; Q3:  $F(5, 403) = 12.14$ ,  $p < .001$ ,  $\eta^2 = .04$ ). Despite reaching statistical significance, also acknowledging the sample sizes' impact, the practical differences in mean scores between the schools were quite varied. Following Cohen's (1988) classifications, eta squared ( $\eta^2$ ), as a measure of effect size, ranged from relatively small (.03) to a maximum of medium effect (.07). *Post hoc* comparisons using the Games–Howell procedure indicated that none of the six schools were significantly different when analysed across the dependent variables all together.

**Artikkel 2:**

**Exploring the relationship between student grades and assessment for learning in Norwegian physical education**

*European Physical Education Review* (2015). Elektronisk publisert før trykk 23.9.2015, s. 1–17, DOI: 10.1177/1356336X15606473.





# Exploring the relationship between student grades and assessment for learning in Norwegian physical education

**Petter E Leirhaug**

Norwegian School of Sport Sciences, Oslo, Norway

European Physical Education Review  
1–17

© The Author(s) 2015

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1356336X15606473

epe.sagepub.com



## Abstract

The purpose of this study was to explore the relationship between student grades and assessment for learning (AfL) in physical education. In educational literature, the focus on formative assessment has grown dramatically, partly because research indicates that good AfL is one of the most effective instructional tools to drive student learning forward. Therefore, a positive influence of AfL on students' grades was hypothesised. The study employed a questionnaire of 1454 PE students (aged 15–19 years) from six upper secondary schools in Norway. Based on theory and factor analysis, a construct to map the experience with AfL was identified. A regression model was used to analyse the relationship between the construct and students' grades, demonstrating that AfL is negatively associated with student grades. While the main finding is statistically small, the study also discusses the influence of other variables, possible explanations for the results and related implications.

## Keywords

Assessment for learning, grading, formative assessment, summative assessment, physical education, Norway

## Introduction

Over the last 20 years in educational research, documentation about the relationship between assessment and learning has proliferated, particularly documentation related to how classroom assessments for formative purposes carry considerable potential to enhance student learning and

---

### Corresponding author:

Petter E Leirhaug, Norwegian School of Sport Sciences, P.O. box 4014, Ullevål Stadion, 0806 Oslo, Norway.

Email: petter.erik.leirhaug@hisf.no

improve teaching (Black and Wiliam, 1998a, 1998b; Gardner, 2012; Hattie, 2009; Hattie and Timperley, 2007; Klenowski and Wyatt-Smith, 2014; Sadler, 1998). When putting formative assessment into practice in schools, the preferred approach is often assessment *for* learning (AfL) (Pryor and Crossouard, 2008). AfL has been defined as ‘... the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there’ (Assessment Reform Group, 2002: 2). Theoretical development and debates on AfL have constantly evolved (Carless, 2010; Gardner, 2012), and have included challenges to some of the empirical evidence of its powerful and positive effect on student achievement (Bennet, 2011; Dunn and Mulvenon, 2009). Nevertheless, most of the literature supports the pedagogical efficacy of several key strategies that characterise AfL practices. Four such strategies, adopted and modified after Black and William (1998a, 1998b) and MacPhail and Halbert (2010), are used to identify AfL in the study reported in this article: (i) sharing learning intentions with students; (ii) sharing criteria of success; (iii) involving students in assessing their own and other students’ learning; and (iv) providing feedback that helps learners advance. These AfL strategies provide teachers with a mechanism for understanding and providing assessment that supports student learning on a daily basis, in contrast to traditional ‘assessment *of* learning’, which typically occurs at the end of a teaching unit and often takes the form of measuring and grading student achievement.

The potential to enhance student learning and strengthen teaching strategies through implementation of AfL is increasingly recognised and advocated in the physical education (PE) literature as well (Georgakis and Wilson, 2012; Green, 2008; Hay, 2006; Hay and Penney, 2013; MacPhail and Halbert, 2010). Up until now, however, little research has been done on the relationship between this relatively new emphasis on formative assessment practices in PE and more traditional end-of-year summative assessments, high-stakes or not, also coupled with the demands of education authorities for accountability evidence. Green (2008) has noted the challenge of promoting formative practices such as AfL in the UK context, where summative practice of examinable PE has become more popular. Thorburn (2008), in presenting a theoretical outline of how a methodology informed by the phenomenology of Merleau-Ponty might improve high-stakes PE examinations, noted the paucity of detailed advice on teaching and showed how different forms of assessment and examination can have serious consequences for subject understanding in PE and even judgements on its overall worthiness. Dyson (2014: 150), in a recent commentary on quality teaching in PE in the USA, warned of the danger of narrowing the focus of PE in a quest for accountability, and stressed the ‘need to create assessments focused on the learning process’.

In Norway, the latest educational reform, introduced in 2006, mandated that schools modify the prevailing ‘assessment *of* learning’ culture and practice by implementing AfL strategies that place the student at the centre of assessment and learning processes (Utdanningsdirektoratet, 2011). The new regulations and overall assessment objectives were directly influenced by the research of, among others, Black and Wiliam (1998a, 1998b). Norwegian teachers are expected to share learning goals and assessment criteria with their students, and all students are entitled to be involved in assessment processes and to receive continuous feedback (‘*underveisvurdering*’ in Norwegian) that promotes learning and supports the development of competence (Tveit, 2014). However, evaluation research has demonstrated that Norwegian schools still have to deal with challenges in integrating assessment as a learning tool in their everyday practice (Møller et al., 2009), and that the school system is riddled with unresolved tensions between the purposes of formative and summative assessments (Tveit, 2014).

While Hay and Penney (2013: 7) view summative assessment 'as a task or programmes that provide or contribute to an account of learning', Harlen (2012) argues that it is important to maintain a distinction between the purposes of formative and summative assessments in the present educational environment. Norway, which does not impose constraints such as final examinations in PE and gives its individual PE teachers primary responsibility for assessment for both formative and summative purposes, offers an excellent context for research on their interrelationship.

Existing research on assessment in PE has primarily focused on teachers' perspectives and methods, with little attention to student perceptions (Chan et al., 2011; Hay and Penney, 2013; Zhu, 2015). Using data derived from a questionnaire administered to students aged 15–19 in six Norwegian upper secondary schools, this study aimed to help fill this gap in the literature and to explore the relationship between the (summative) grades the students received in PE and (the formative practice of) AfL in PE.

### *Research on assessment and the research hypothesis*

There is a general expectation that grades should objectively reflect the learner's competence (Klapp Lekholm, 2008). Yet researchers, both in Scandinavia and internationally, have for years demonstrated otherwise (McMillan, 2003; Svennberg et al., 2014; Tveit, 2014). With regard to grading in PE, teachers sometimes emphasise and assign grades based at least in part on criteria completely different from the ones found in the official curriculum and assessment regulations (Redelius and Hay, 2009; Young, 2011). The tradition has been to give attention to administrative issues, such as social behaviour, punctuality and dress, when assessing students (Hay and Penney, 2013; van der Mars and Harvey, 2010). Recent reports indicate that students are still unclear of assessment criteria and on how their grade in PE is determined (Redelius and Hay, 2012; Zhu, 2015). Furthermore, researchers have found that PE teachers do not always grade the way they think they do (Chatzopoulos et al., 2006), and that they have difficulties accounting for their own assessment practices and grading criteria (Annerstedt and Larsson, 2010; Hay and MacDonald, 2008). In addition, different teachers express that they emphasise the same elements, for example effort or motor skills, in very different degrees (Vinje, 2008). Melograno (2007) has demonstrated that an emphasis on different criteria can lead to radically different judgements and grades for the same student performance. Such findings, of course, pose serious threats to the validity and reliability of assessment and grading in PE (Annerstedt, 2010). Yet when Norwegian PE teachers are asked about fairness, their primary concern is the comparability of grading practices among schools, rather than the validity of their own judgements (Prøitz and Borgen, 2010).

Historically, there has been a generally sceptical attitude towards assessment and grading among PE teachers, and many have argued that assessment and grading are somewhat contradictory to the aim of stimulating the individual to adopt and maintain an active lifestyle after finishing school (Carrol, 1994; Hay and Penney, 2013; Matenin and Tannehill, 1994). The dubious attitude observed is most likely to be understood in light of traditional measurement conceptions underpinned by paradigms of scientific measurement, social efficiency and behaviourist learning theories, where assessment remains distinct from pedagogy. This fits well with the view that the term 'assessment' has been 'commonly used by teachers in reference to grades' (Green, 2008: 79), and is typically associated with summative end-of-term or end-of-year reports on each student's achievement level. This approach has been challenged by the social constructivist paradigms of knowledge that emerged in the late 20th century, which emphasise the interrelatedness of

assessment, curriculum and learning theory, and place the learner at the centre of the learning process (Pryor and Crossouard, 2008). These socio-culturally informed learning theories have stimulated the development of AfL (Gardner, 2012; Hay, 2006), which Hay and Penney (2013: 110) see

... as a tool that can be utilised to generate discussions and new thinking about learning and learning opportunities in PE within and beyond schools. It is concerned with quality and equitable learning opportunities and experiences now and in the future, for all students.

The few studies on PE and assessment conducted since the 2006 reform suggest that most Norwegian PE teachers find other challenges in student assessment more urgent than developing and implementing effective AfL (Arnesen et al., 2013; Engvik, 2010; Prøitz and Borgen, 2010; Vinje, 2008). From a validity perspective, assessment and grading in Norwegian PE have been described as 'a minefield of conflicting guidelines and intentions' (Sandvik et al., 2012: 153; translation by author). Moreover, Leirhaug and MacPhail (2015) have demonstrated that even PE teachers who use the term AfL do not necessarily understand it well enough to use it effectively. Nevertheless, based on research evidence of AfL's positive effect on student achievement (Gardner, 2012; MacPhail and Halbert, 2010), even partial adoption might enhance the learning outcomes of students who report having more AfL experiences. This is the basis for the hypothesis (H1) in the current study:

There is a positive relationship between student-reported higher levels of AfL experiences in PE and better grades in the subject.

The socio-cultural perspective underpinning AfL also reminds us that there are a host of other variables acting on school performance. For example, gender, immigrant background, socioeconomic background and parental education are variables that correlate with Norwegian students' school results (Dale, 2008; Steffensen and Ziade, 2008). To contextualise the use of student grades as the dependent variable in this study, the next section begins with a glance into the system for assessment and grading in Norwegian PE.

## Methodology

### *Research context and the dependent variable*

Norway mandates 10 years of compulsory school beginning at age six, and gives all children the right to attend upper secondary school for an additional three years. PE is a compulsory subject throughout, and there is one common national curriculum structured into main subject areas with specified 'competence aims'. In upper secondary school, the competence aims describe expected learning outcomes in three areas; *Sports and dance*, *Outdoor life (Friluftsliv)* and *Exercise and lifestyle*. The competence aims constitute the basis for all student assessments (summative and formative) in PE, and thus the grading.

The Norwegian grading scale in all subjects is 1–6, where 1 represents very low competence and 6 represents excellent competence. In this study, the last received PE grade reported by the student is the dependent variable, and is referred to as the *student grade*. In the current study, this ordinal variable is treated as continuous. This is a common practice in educational research (Muijs, 2011), although one might question whether the distance between numbers is equal in a

grading scale. In Norway, there is also an option of ‘no grade in the subject’, due to difficulties in assessing some students. Those students who reported receiving no grade ( $n = 32$ ) were excluded from the study.

Along with other grades, those awarded in PE at the end of compulsory school and after upper secondary school can be considered high-stakes assessments: they are used as a selection instrument for the next level in the school system and by employers. The national authorities’ responsibility for high-stakes assessment has traditionally been limited to the examination system, but like in many countries around the world (Pühse and Gerber, 2005) there is no final exam for PE in Norway. Consequently, and similar to the expected formative assessment practices, PE teachers have sole responsibility for determining these high-stake grades. Traditionally, teachers have based them to a great extent on continuous observations and assessments (Engvik, 2010; Prøitz and Borgen, 2010).

### *Participants and data collection*

The participants in this study were 1454 students (820 girls, 634 boys; difference due to the vocational subjects offered at the schools), mainly aged 15–19 years, from six upper secondary schools in Norway ( $N = 2046$ , response rate 71.1%). The six schools were recruited in a convenience sample (Bryman, 2012) for a research project studying AfL in PE from different perspectives. In addition to solicitation of student data through the questionnaire, the project included focus group interviews with all PE teachers at each of the selected schools.

The questionnaire was intended to capture a broad picture of the students’ perspectives on PE, with a focus on assessment practice, and developed specifically for this project. As such, it was not a pre-validated instrument, and the results should be regarded as preliminary. In addition to AfL theory, the development was informed by questions previously posed to PE students in Sweden (Larsson and Redelius, 2004; Redelius and Hay, 2012). The questionnaire also included questions about student background and participation in movement activity outside school. All of the questions were formulated to fit an online survey format and piloted with three different PE classes, totalling to 67 students. The results of the pilot generated only minor modifications to a few questions. The final online questionnaire had 32 main questions, and took 15–20 minutes to answer.

In all six schools, the questionnaire was conducted during an ordinary lesson. In four of them, the researcher himself visited the classrooms, presented the study and encouraged honest responses. The time schedule in the other two schools required classes to fill out the questionnaire at the same time; therefore, it was administered by the regular teacher after reading an information letter from the researcher. At four schools, the students had personal laptops, which they used for the questionnaire; in the other two schools, the students had to move to a computer lab. All data for the project were collected in 2011 and 2012. The study was approved by the Norwegian Social Sciences Data Service (NSD), and students were assured of anonymity and informed of their right to withdraw from the study at any time.

### *Development of independent variables*

Data were initially screened for outliers as well as missing and implausible values. Nine independent variables anticipated to influence grades were obtained more or less directly out of the questionnaire: *gender of student, immigrant background, disability, parental education, weekly*

*physical activity level, participation in competitive sports, gender of teacher (in PE)*, if the student ranked *PE more important* than other subjects and *physical activity level in PE*. Most of these variables have been previously shown by Peev (2001) as important when looking at the assessment practices of Norwegian PE teachers, and several of them draw attention to the challenges of equal and socially just teaching and learning in PE (Green, 2008; Tinning, 2010). Nominal variables were encoded as dummy variables to enter analysis. *Weekly physical activity level*, measured on a five-point scale from ‘every day’ to ‘very rare’, was treated as a continuous variable in accordance with the logical increase in activity level, although there is no real measurable distance between the scale points. This variable showed kurtosis and skewness in an acceptable range and met the normality criteria.

Since the questionnaire did not entail pre-validated constructs, theoretical assumptions were used to elicit a total of 31 items (overview in the Appendix) deemed relevant for identifying students’ experiences with AfL strategies, their perceptions of grading practice and their motivation to perform in PE and in school in general. All items were arranged to gather students’ responses on a five-point Likert scale spanning from ‘Totally disagree’ to ‘Totally agree’. Since the instrument had not previously been validated, explorative factor analyses were chosen to reveal relationships in the data (Muijs, 2011; Ringdal, 2013). Five items showed communalities below .5; three of them were excluded. The other two items, ‘the students are engaged in peer-assessment’ and ‘the competence aims are the basis for assessment in PE at our school’, were retained according to the theoretical importance for AfL.

A principal axis factor analysis was then conducted on the remaining 28 items. This involved direct oblimin rotation, because factors did show correlations. The Kaiser–Meyer–Olkin measure verified the sampling adequacy ( $KMO = .89$ ), and all items were well above the limit of .5 (Field, 2013). The Kaiser criterion of eigenvalue  $>1$  suggested four factors. These four factors explained 60.52% of the variance; items with loadings are presented in Table 1. Cronbach’s alpha coefficients, also reported in Table 1, were all greater than .70, which indicates satisfactory internal reliability (Field, 2013; Ringdal, 2013).

The first factor in Table 1 consists of six items in which the four key AfL strategies cited in the introduction can be identified. Subsequently, this construct was found to reflect students’ AfL experience, and provided the study with a main independent variable, *AfL perception*. The analysis revealed that the items clustered on the three other factors were also theoretically meaningful. They were assigned the names *teacher’s grading, mastery and motivation* and *well-being*. From general educational research, factor number two recognises the teacher as one of the most important contributors to student learning (Hattie, 2009; Nordenbo et al., 2008), here with a particular focus on the teacher’s grading. Trouilloud et al. (2002) studied how teachers’ expectations of students influence their way of grading in PE. They found connections between teachers’ expectations and students’ achievements and described how teachers’ conceptions of students can lead to self-fulfilling prophecies with consequences for students’ achievement and hence grades.

The third factor appears as a cluster of items asking about enjoyment and motivation in PE as it relates to experienced mastery of subject-related tasks. Ideally, students should encounter challenges that they are familiar with and believe they can handle: ‘This is something I can manage’ (Smith, 2009: 31). The importance of *mastery and motivation* is emphasised in a variety of pedagogical theories. Self-determination theory (Ryan and Deci, 2000), for example, asserts that perceived competence, autonomous motivation and action orientation are positively associated with better performance in a given subject and hence with student grades. The fourth and last factor

**Table 1.** Summary of exploratory factor analysis (principal axis factoring with direct oblimin rotation) based on 28 items elicited from a questionnaire concerning physical education and assessment.

Item	Variables and factor loadings			
	AfL perception	Mastery and motivation	Well-being	Teacher's grading
The physical education teacher has explained assessment for learning.	<b>.78</b>	-.06	-.11	.01
I receive feedback on where I am in the learning process.	<b>.59</b>	-.02	-.01	-.01
I am being engaged in self-assessment.	<b>.63</b>	.01	.13	.03
The competence aims are the basis for assessment in physical education at our school.	<b>.49</b>	.06	.13	.03
I know what to do to receive a better grade.	<b>.59</b>	-.01	-.14	.13
The students are engaged in peer assessment.	<b>.43</b>	.09	.15	-.07
The physical education teacher assigns fair grades.	.22	-.07	-.07	<b>.66</b>
I think I receive the grade which I deserve.	.02	-.02	.06	<b>.89</b>
I often experience mastery in physical education.	-.01	<b>.67</b>	-.23	.05
I am motivated to exercise and learn new skills.	.03	<b>.65</b>	-.11	-.03
I have fun in almost all lessons.	.08	<b>.40</b>	-.20	.19
I am satisfied with my achievements in physical education.	-.04	<b>.62</b>	-.11	.10
I enjoy performing in physical education.	.11	<b>.70</b>	.17	-.09
I feel well in school.	.02	.08	<b>-.81</b>	-.01
I have good friends in class.	-.01	.14	<b>-.72</b>	-.03
Eigenvalue (estimated by sum of squared coefficient of factors in the structure matrix in SPSS.)	2.23	2.24	1.29	1.34
Cronbach's $\alpha$	.78	.80	.82	.79

Total variance explained = 60.5%

Note: Factor loadings reported from pattern matrix in SPSS. What to report when rotation is used is discussed in the methodological literature; both Field (2013) and Muijs (2011) suggest that pattern matrix is most common. Factor loadings over .40 appear in bold. Correlation between factors: AfL/Teacher's grading = .397; AfL/Mastery and motivation = -.165; AfL/Well-being = .431; Teacher's grading/Mastery and motivation = -.341; Teacher's grading/Well-being = .233; Mastery and motivation/Well-being = -.372

AfL: assessment for learning.

revealed in the analysis, general *well-being* at school, can be assumed to be a possible influencing variable.

### Analyses

Variables were checked for assumptions, and primary correlations and regression analyses were conducted to assess associations among all variables, particularly associations between the dependent variable *student grade* and the 13 independent variables. Based on the primary analysis, variables that did not contribute significantly were excluded and the best-fit linear model to predict student grade in PE was chosen. Multiple regression with forced entry (all variables enter the multiple regression simultaneously) was evaluated to give enough relevant information, keeping in mind the exploratory aim of the study and the somewhat preliminary constructs used in it. All analyses were performed using IBM SPSS version 21.0 software.

**Table 2.** Regression table of nine variables influencing *student grade* in physical education (95% bias corrected and accelerated confidence intervals reported in parentheses; confidence intervals based on 1000 bootstrap samples).

	B	SE B	Beta ( $\beta$ )	t	Sig. (p)
Constant	1.78	0.12		16.59	$p < .001$
AfL perception	-0.03 (-0.04, -0.02)	0.01	-.14	-5.83	$p < .001$
Teacher's grading	0.14 (0.12, 0.16)	0.01	.34	14.81	$p < .001$
Mastery and motivation	0.05 (0.3, 0.6)	0.01	.21	8.16	$p < .001$
Gender of student	0.22 (0.15, 0.30)	0.04	.12	5.49	$p < .001$
Gender of teacher	-0.15 (-0.23, -0.07)	0.05	-.07	-3.49	$p < .001$
Weekly physical activity level	0.14 (0.11, 0.18)	0.02	.18	8.35	$p < .001$
Physical activity level in PE	0.49 (0.37, 0.61)	0.06	.21	8.98	$p < .001$
Immigrant background	0.11 (0.00, 0.22)	0.06	.04	2.07	$p = .039$
Parental education	0.12 (0.08, 0.16)	0.02	.14	6.59	$p < .001$

Adjusted  $R^2 = .41$

$F = 112.9$ , sig.  $F$  change:  $p < .001$

AfL: assessment for learning.

## Results

A first regression analysis with the construct *AfL perception* as the only predictor revealed a statistically significant relationship, but accounted for only 2% of the variance in the dependent variable *student grade* ( $R^2 = .02$ ). Each of the other independent variables was similarly analysed; there were no statistically significant relationships between *student grade* and the independent variables *well-being*, *disability*, *PE more important* or *participation in competitive sports*. These variables were removed. Even though the variables *immigrant background* and *gender of teacher* showed only small statistical contributions – perhaps due to the impact of the sample size on correlations – they were retained due to their potential importance in future research.

Table 2 provides the results of the regression analysis encompassing nine independent variables. The chosen model accounts for about 41% of the variance in *student grade* ( $R^2 = .41$ ), which according to Muijs (2011: 145) indicates a model of ‘moderate fit’. The  $F$ -ratio is highly significant ( $F = 112.9$ ,  $p < .001$ ) and analyses of linearity, homoscedasticity, multicollinearity and residuals did not show values that violate the assumptions (Field, 2013; Ringdal, 2013). To ensure the assumptions, the confidence intervals are reported based on 1000 bootstrap samples.

The results presented in Table 2 illustrate how the main research interest variable, *AfL perception*, had a negative although statistically small influence on *student grade* (standardised  $\beta = -.14$ ). A comparison of all the standardised betas in Table 2 reveals that the variable *teacher's*



*grading* elicited the strongest influence on *student grade*. Together with *physical activity level in PE* and *mastery and motivation*, which also have relatively high betas, these three variables alone accounted for about 31% of the variance in *student grade*. For the other five independent variables, the model showed smaller but significant contributions.

## Discussion

Informed by AfL theory and research on feedback (Gardner, 2012; Hattie, 2009; Hattie and Timperley, 2007), this study hypothesised that students who reported more experience with AfL in PE would receive better grades in the subject. The hypothesis (H1) anticipated that AfL would have a positive effect on grades, but the results indicated otherwise. Indeed, the regression analysis showed that the relationship between a reported higher level of AfL and student grades in upper secondary school PE in Norway was slightly negative (standardised  $\beta = -.14$ ), although reported AfL experience appears to account for under 2% of the variance in the dependent variable. Given the large sample size, the statistical significance of such minor results can be called into question. However, this does not affect the interpretation: in this study a finding of no statistical relationship would also lead to rejection of H1.

In other words, students who reported more experience with AfL in PE classes did not, on average, receive better grades. However, this does not necessarily imply that AfL did not improve their learning outcomes. In his Norwegian Master's thesis on AfL in PE, Eide (2011) observed that innovative AfL implementation led to a decrease in the mean student grade, but suggested that one reason for this may be that both teachers and students have become more assessment literate and aware of the criteria for success in the subject. Even so, this would only apply when AfL is already embedded in pedagogical practice, which is probably not the case in most schools. Recalling a lack of deeper understanding even among PE teachers who referred to AfL as what they were doing, or attempting to do, in their assessment practice (Leirhaug and MacPhail, 2015), it is more likely that the initiated major change in assessment has not yet succeeded, partly because 'teachers do not hold the essential competence to implement AfL in their teaching' (Smith, 2011: 60). This indicates that to master this competency, the average PE teacher will probably need extensive support as well as time to engage in long-term professional development.

A complementary explanation of the finding that student-reported higher level AfL experiences do not lead to better grades is that the learning outcomes supported by formative assessments can be different from the criteria that teachers actually use when they assign grades to students. This interpretation is well supported by previous research on assessment in PE, suggesting that many teachers have their own ways of grading, and that the criteria used are not necessarily consistent with either the curriculum or with what PE teachers managed to make transparent to their students (for a Swedish perspective, see Annerstedt and Larsson, 2010; Redelius and Hay, 2009, 2012; Svennberg et al., 2014). If this were the case, it indicates not only a misalignment between formative and summative assessment practices and highlights the need to parallel PE assessment criteria with both instructional practices and curricular goals, but also shows how issues involved in AfL are intimately related to more general discussions of quality teaching, teacher philosophies, content knowledge and what is to be valued in the name of PE (Dyson, 2014; Green, 2008; Tinning, 2010).

Considering that the main finding on the construct *AfL perception* only explained between 1% and 2% of the variance in students' grades, it becomes interesting to look into the influence of the other variables. *Teacher's grading* was the variable with the strongest influence on the students'

grades in PE. Students who reported that the teacher were fair in grading and student assessment was invariably among those who received higher grades. That the teacher has such significance is consistent with previous research (Hattie, 2009). Particularly in a school system such as Norway's, in which assessment is decentralised, this places the responsibility for distribution and negotiation of power relations in the hands of individual schools and teachers, which many teachers do not realise (Palm, 2014). Students that report fair grading are most likely to be the most assessment literate, those who already are able to 'read the teacher's clues' regarding behaviours that will be rewarded with better grades, and orientate their actions accordingly (Pryor and Crossouard, 2008). Both in regard to equity and assessment validity, this becomes problematic when PE teachers' grading practices are inconsistent with the formal curriculum goals and they continue to refer to a 'gut-feeling' when asked to explain student grades (Annerstedt and Larsson, 2010; Svennberg et al., 2014). On the other hand, the discrepancy between actual grading criteria and what PE teachers emphasise in the classroom as important to learn in PE may partially explain why a higher level AfL experience was not positively related to better grades in this study.

The variables *physical activity level in PE* and *weekly physical activity level* also appeared relatively important in the model, indicating that students more active in PE lessons and out of school receive better grades than other students who reported less movement activity. As Green (2008) and Tinning (2010) have observed, this makes intuitive sense, as being physically active is an inherent part of PE. This finding is also consistent with the view that PE has traditionally encouraged and assessed activity (level) rather than learning (Hay and Penney, 2013), and Ericsson (2011) has previously shown that extended physical activity within school correlates with higher grades. Furthermore, recalling the items of the construct *mastery and motivation*, it is no surprise that students who enjoy performing in PE, and who report frequent experiences of mastery, also receive higher grades. Looking more closely into the subject, recent research suggests that motivation is closely connected to the content and methods used in concrete teaching and learning situations (Ding et al., 2013). Jakkola et al. (2013: 281) have, for example, 'revealed that amotivation during fitness testing was at a higher level than in [the] general PE program'. Again, observations such as these draw attention to the significant influence that educational context and curriculum have on the content validity of assessment and the advocacy of quality assessment through alignment between curriculum, pedagogy and assessment (Klenowski and Wyatt-Smith, 2014; Penney et al., 2009).

Regarding the remaining variables, the analysis revealed they made only minor contributions to the dependent variable. The standardised beta values for *parental education* ( $\beta = .14$ ) and *gender of student* ( $\beta = .12$ ) appeared to have some, although small, statistical importance in the model. Similar to most Western countries, the boys in this study received better grades in PE on average than the girls (Hay and Penney, 2013; Pühse and Gerber, 2005). In Norway, this is the opposite to all other compulsory school subjects, where girls earn better grades on average than boys (Klomsten, 2013; Steffensen and Ziade, 2008). In general, students reporting parents with higher education receive better grades (Dale, 2008); the study found this to be true in PE as well. The results also indicate that students born and raised in Norway receive slightly higher grades than students who report an immigrant background, and the negative beta on *gender of teacher* tells us that students with a male teacher will receive better grades on average. The negative value is due to an expectation of this outcome to be the other way around, based on earlier findings by Peev (2001).

This study has many limitations, including those that always accompany the use of self-reported questionnaires, and the fact that a cross-sectional study of six schools has unknown but limited

representativeness (Bryman, 2012). Another concern could be the construct used to map student reports of AfL. However, based on six items that were statistically supported in factor analysis and theoretically cover the four key AfL strategies presented earlier, it is possible to argue the opposite: that the study presents an AfL construct in PE that can be utilised and validated in future research. Considering the explorative nature of this study, and also acknowledging the use of an instrument that was not pre-validated, regression analyses were performed only with forced entry. The model did not change much by the introduction of product variables, which is why they were not included.

Returning to the main finding, which indicated that the effect of AfL on grades is negative rather than positive, with the Norwegian context as an example, one major challenge in PE appears to be finding beneficial ways of aligning assessments with summative and formative purposes. Berry (2010) noted that even teachers who are motivated to apply AfL strategies find it rather difficult to fight the traditional assessment culture and the pressures of grading and accountability. Questions about how to understand and enact AfL in PE, as well as how it should be valued in the summative practice of grading, remain unanswered. Genuine and effective AfL implementation appears to be challenged from a school system stressing accountability at the political level, the PE teachers' implicit or explicit understanding and communication of the subject's most valuable content, and a traditional PE assessment culture that has focused on summative assessments and student grading. The results from this study support the call of Georgakis and Wilson (2012) for formative assessment practices that can be used to inform teaching and drive student learning in PE, coupled with a 'shift from poor quality assessment of learning to more productive *assessment of learning*' (Georgakis and Wilson, 2012: 50, italics in original).

## Conclusion

This study was undertaken to explore the relationship between student grades and AfL in PE, and obtained a result that contradicted the expectation of hypothesis H1. A student-reported higher level of AfL experiences, here analysed as a construct based on six questionnaire items and explored in a regression model, appears negatively associated with better grades. The association was minimal and explained less than 2% of the variance in grades. Based on the standardised beta values illustrated in Table 2, it is tempting to conclude that in Norwegian upper secondary school PE, the construct AfL perception contributes about as much to student grades as being a girl rather than a boy. In Norway, as in most other Western countries, girls earn slightly lower grades than boys in PE on average (Pühse and Gerber, 2005; Steffensen and Ziade, 2008). However, in contrast to the gender issue, where it is an educational goal to reduce this inequity and eliminate an imbalance, the purpose of introducing AfL in PE is to raise student achievement and, hence, on average, improve all student grades.

Although this study was conducted as a preliminary exploration solely in a Norwegian context, the minor, and indeed negative, relationship between AfL and grades that it found represents a major challenge for education authorities promoting formative assessment practices, as well as for PE teachers and assessment researchers on how to analyse and improve the quality of assessment work in general and grading in particular. In the worst-case scenario, if PE pedagogical practice and future research cannot demonstrate a better correspondence between the formative practice of AfL and the summative practice of grading, it is likely that most students will not find it worthwhile to engage with AfL in PE. Considering the strong student-focused approach to pedagogy that underpins AfL, this, in turn, will make genuine implementation of AfL nearly impossible.

### Appendix: Overview of items elicited from student questionnaire to be included in factor analysis. Norwegian original and English translation

Included statements/questions (Norwegian original)	Translation	Inverted scale = X
Main question: Hvor godt passer disse påstandene med din oppfatning av karakterer og vurdering i kroppsøving? [How well do these statements match your perception of grades and assessment in PE?] Items:		
1 Kroppsøvlingslærer samtaler med meg om karakteren	The PE teacher discusses my grade with me.	
2 Jeg synes jeg får den karakter jeg fortjener	I think I receive the grade which I deserve.	
3 Elevene er med og vurderer hverandre	The students are engaged in peer assessment.	
4 Jeg er fornøgd med kroppsøvlingskarakteren min	I am happy with my grade in physical education.	
5 Kroppsøvlingslæreren gir rettferdige karakterer	The physical education teacher assigns fair grades.	
6 Vurdering handler mer om hva jeg skal lære enn om karakter	Assessment is more about learning than grading.	
7 Jeg vet hva som kreves for å få de ulike karakterene	I know what I have to do to earn the different grades.	
8 Kroppsøvlingslæreren har forklart om vurdering for læring	The physical education teacher has explained assessment for learning.	
9 Jeg får være med å vurdere meg selv	I am being engaged in self-assessment.	
10 Jeg vet hva jeg må gjøre for å få bedre karakter	I know what to do to receive a better grade.	
11 Ved vår skole er kompetansemålene grunnlaget for vurdering i kroppsøving	The competence aims are the basis for assessment in physical education at our school.	
Main question: Hvordan passer disse påstandene med din selvoppfatning? [How well do these statements fit your self-perception?] Items:		
12 Jeg trives godt på skolen	I feel well in school.	
13 Jeg opplever ofte mestring i kroppsøving	I often experience mastery in physical education.	
14 Jeg har gode venner i klassen	I have good friends in class.	
15 Jeg er fornøyd med mine prestasjoner i kroppsøving	I am satisfied with my achievements in physical education.	
16 Jeg føler meg ofte utrygg i kroppsøving	I often feel unsafe in physical education.	X
17 Jeg er generelt skolelei	I am tired of school.	X

(continued)

**Appendix** (continued)

Included statements/questions (Norwegian original)	Translation	Inverted scale = X
18 Kroppsøving er det dårligste faget mitt	PE is the subject in which I perform the worst.	X
19 Jeg er motivert for å trene og lære nye ferdigheter	I am motivated to exercise and learn new skills.	
20 Jeg liker å vise meg fram i kroppsøving	I enjoy performing in physical education.	
21 Jeg liker å trene/stå på så jeg blir skikkelig fysisk sliten	I enjoy exercising until I am exhausted.	
Main question: Hvor godt passer disse påstandene med din oppfatning av kroppsøving? [How well do these statements match your perception of PE?] Items:		
22 Jeg har det gøy i nesten alle timene	I have fun in almost all lessons.	
23 Jeg skulle helst sluppet å være med	I would rather not participate in PE.	X
24 Jeg blir aldri sliten i timene	I never become exhausted in PE class.	X
Main question: Hvor godt passer disse påstandene med din oppfatning av undervisningen i kroppsøving? [How well do these statements match your perception of the teaching in PE?] Items:		
25 Jeg får hjelp når jeg trenger det	I receive help with learning when I need it.	
26 Jeg får tilbakemeldinger om hvor jeg er i læreprosessen	I receive feedback on where I am in the learning process.	
27 Jeg blir ofte nødt til å gjøre ting jeg ikke liker	I often have to complete tasks that I do not enjoy.	X
28 Læreren har god oversikt over hva jeg kan og ikke kan	The teacher knows of what I can and cannot do.	
29 Lærer forskjellsbehandler elever	The teacher treats students differently.	X
30 Lærer er flink til å motivere og engasjere elevene	The teacher excels at motivating and engaging students.	
31 Egenvurdering brukes systematisk	Self-assessment is used systematically.	

**Acknowledgements**

Thanks go to Claes Annerstedt, Kari Smith and Trond Solhaug for their advice at different stages of development of this manuscript. Thanks also go to two anonymous reviewers who offered insight and suggestions that undoubtedly improved the manuscript.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship and/or publication of this article.

## References

- Annerstedt C (2010) Karaktersetting i kroppsøving: Problematisk, urettferdig og neppe likeverdig [Grading in physical education: Problematic, unfair and probably inequal]. In: Steinholt K and Gurholt KP (eds) *Aktive liv: Idrettspedagogiske perspektiver på kropp, bevegelse og dannelse*. Trondheim: Tapir, pp.233–254.
- Annerstedt C and Larsson S (2010) 'I have my own picture of what the demands are . . .' Grading in Swedish PEH – problems of validity, comparability and fairness. *European Physical Education Review* 16(2): 97–115.
- Annesen TE, Nilsen AK and Leirhaug PE (2013) 'Den læreplanen som ikkje kan tilpassast mi undervisning, finst ikkje.' Vurdering og undervisning i kroppsøving etter Kunnskapsløftet [Assessment and teaching in physical education after the Knowledge Promotion Act 2006]. *Tidsskriftet FOU i praksis* 7(3): 9–32.
- Assessment Reform Group (2002) *Assessment for learning: 10 principles*. Available at: [http://assessmentreformgroup.files.wordpress.com/2012/01/10principles\\_english.pdf](http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf) (accessed 14 December 2012).
- Bennet RE (2011) Formative assessment: a critical review. *Assessment in Education: Principles, Policy and Practice* 18(1): 5–25.
- Berry R (2010) Teachers' orientations towards selecting assessment strategies. *New Horizons in Education* 58(1): 96–107.
- Black P and Wiliam D (1998a) Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5(1): 7–73.
- Black P and Wiliam D (1998b) Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80(2): 139–148.
- Bryman A (2012) *Social research methods*. 4th ed. Oxford, New York: Oxford University Press.
- Carless D (2010) Prospects of implementing assessment for learning. *Assessment in Education: Principles, Policy and Practice* 12(1): 39–54.
- Carrol B (1994) *Assessment in Physical Education: A Teacher's Guide to the Issues*. London: Falmer.
- Chan K, Hay P and Tinning R (2011) Understanding the pedagogic discourse of assessment in Physical Education. *Asia-Pacific Journal of Health, Sport and Physical Education* 2(1): 3–19.
- Chatzopoulos D, Erdmann R and Tsormbatzoudis H (2006) Pupils' grading: Do teachers grade according to the way they report? *International Journal of Physical Education* 43(1): 4–10.
- Dale EL (2008). *Fellesskolen – Reproduksjon av Sosial Ulikhet* [One School for All – Reproduction of Social Inequality]. Oslo: Cappelen.
- Ding H, Sun H and Chen A (2013) Impact of expectancy-value and situational interest motivation specificity on physical education outcomes. *Journal of Teaching in Physical Education* 32(3): 253–269.
- Dunn KE and Mulvenon SW (2009) A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation* 14(7): 1–11.
- Dyson B (2014) Quality physical education: A commentary on effective physical education teaching. *Research Quarterly for Exercise and Sport* 85(2): 144–152.
- Eide LH (2011) *Vurdering for læring i kroppsøving* [Assessment for learning in physical education]. Master's Thesis, Norwegian School of Sport Sciences, Oslo, Norway.
- Engvik G (2010) Elevvurdering i faget kroppsøving – Noen utfordringer. [Student assessment in physical education – Some challenges.] In: Dobsen S and Engh R (eds) *Vurdering for Læring i Fag*. Kristiansand: Høyskoleforlaget, pp.201–213.
- Ericsson I (2011) Effects of increased physical activity on motor skills and marks in physical education: an intervention study in school years 1 through 9 in Sweden. *Physical Education and Sport Pedagogy* 16(3): 313–329.
- Field A (2013). *Discovering statistics using IBM SPSS statistics*. 4th ed. London: Sage.
- Gardner J (ed) (2012) *Assessment and Learning*. 2nd ed. London: Sage.

- Georgakis S and Wilson R (2012) Australian physical education and school sport: An exploration into contemporary assessment. *Asian Journal of Exercise & Sport Science* 9(1): 37–52.
- Green K (2008) *Understanding Physical Education*. London: Sage.
- Harlen W (2012) On the relationship between assessment for formative and summative purposes. In: Gardner J (ed) *Assessment and Learning*. 2nd ed. London: Sage, pp.87–102.
- Hattie J (2009) *Visible Learning. A Synthesis of over 800 Meta-analysis Relating to Achievement*. London: Routledge.
- Hattie J and Timperley H (2007) The power of feedback. *Review of Educational Research* 77(1): 81–112.
- Hay P (2006) Assessment for learning in physical education. In: Kirk D, Macdonald D and O’Sullivan M (eds) *The Handbook of Physical Education*. London: Sage, pp.312–325.
- Hay P and Macdonald D (2008) (Mis)appropriations of criteria and standards referenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy and Practice* 15(2): 153–168.
- Hay P and Penney D (2013) *Assessment in Physical Education. A Sociocultural Perspective*. New York/Oxon: Routledge.
- Jakkola TT, Sääkslahti A, Yli-Piipari S, et al. (2013) Student motivation associated with fitness testing in the physical education context. *Journal of Teaching in Physical Education* 32(3): 270–286.
- Klapp Lekholm A (2008) Grades and grade assignment: Effects of student and school characteristics. *Göteborg Studies in Educational Sciences, Vol. 269*. Göteborg: Acta Universitatis Gothoburgensis.
- Klenowski V and Wyatt-Smith C (2014) *Assessment for Education: Standards, Judgments and Moderation*. London: Sage.
- Klomsten AT (2013) Hvordan organiseres kroppøvningsfaget i norske skoler: Kjønnsblandet eller kjønnsdelt? [How is the teaching of boys and girls organized in Norwegian physical education?] *Tidsskriftet FOU i praksis* 7(3): 59–82.
- Larsson H and Redelius K (eds) (2004) *Mellan nytta och nöje: bilder av ämnet idrott och hälsa*. [Between Benefit and joy: Pictures of the Subject PE and Health.] Stockholm: GIH.
- Leirhaug PE and MacPhail A (2015) ‘It’s the other assessment that is the key’: three Norwegian physical education teachers’ engagement (or not) with assessment for learning. *Sport, Education and Society* 20(5): 624–640.
- MacPhail A and Halbert J (2010) ‘We had to do intelligent thinking during recent PE’: Students’ and teachers’ experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy and Practice* 17(1): 23–39.
- Matenin M and Tannehill D (1994) Assessment and grading in physical education. *Journal of Teaching in Physical Education* 13(4): 395–405.
- McMillan JH (2003) Understanding and improving teachers’ classroom assessment decision-making: Implications for theory and practice. *Educational Measurement: Issues and practice* 22(4): 34–43.
- Melograno VJ (2007) Grading and report cards for standard-based physical education. *Journal of Physical Education, Recreation and Dance* 78(6): 45–53.
- Muijs D (2011) *Doing Quantitative Research in Education with SPSS*. 2nd ed. London: Sage.
- Møller J, Prøitz TS and Aasen P (eds) (2009) *Kunnskapsløftet – tung bør å bære* [the Knowledge Promotion – an evaluation report]. Report 42/09. Oslo: NIFU STEP.
- Nordenbo SE, Larsen MS, Tiftikçi N, et al. (2008) *Lærerkompetanser og elevers læring i førskole og skole*. [Teacher Competence and Students’ Learning in Schools.] København: Dansk Clearinghouse for Uddannelsesforskning.
- Palm T (2014) *Assessing with Foucault. A critical study of assessment in physical education*. Master’s Thesis, Norwegian School of Sport Sciences, Oslo, Norway.
- Peev OD (2001) *Nasjonal standard i vurderingen – realitet, visjon eller luftspeiling?* [National standard in assessment – reality or not?]. Master’s Thesis, Norwegian School of Sport Sciences, Oslo, Norway.
- Penney D, Brooker R, Hay P, et al. (2009) Curriculum, pedagogy and assessment: Three message systems of Schooling and dimensions of quality physical education. *Sport, Education and Society* 14(4): 421–444.

- Pryor J and Crossouard B (2008) A socio-cultural theorisation of formative assessment. *Oxford Review of Education* 34(1): 1–20.
- Prøitz TS and Borgen JS (2010) *Rettferdig standpunktvurdering – det (u)muliges kunst?* [Fair grading – an (im)possible art?]. Report 16/2010. Oslo: NIFU STEP.
- Pühse U and Gerber M (eds) (2005) *International Comparison of Physical Education: Concepts, Problems, Prospects*. Oxford: Meyer & Meyer Sport.
- Redelius K and Hay P (2009) Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review* 15(3): 275–294.
- Redelius K and Hay P (2012) Student views on criterion-referenced assessment and grading in Swedish physical education. *Physical Education and Sport Pedagogy* 17(2): 211–225.
- Ringdal K (2013) *Enhet og mangfold. Samfunnsvitenskaplig forskning og kvantitativ metode* [Oneness and manifold. Social science and quantitative method]. 3rd ed. Bergen: Fagbokforlaget.
- Ryan RM and Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55(1): 68–78.
- Sadler DR (1998) Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy and Practice* 5(1): 77–85.
- Sandvik LV, Engvik G, Fjørtoft H, et al. (2012) *Vurdering i skolen. Intensjoner og forståelser – Delrapport 1*. [Assessment in the School. Intentions and understandings – Report 1.] Trondheim: NTNU. Available at: <http://www.udir.no/Upload/Rapporter/2012/fivis.pdf?epslanguage=no> (accessed 14 December 2012).
- Smith K (2009) Samspillet mellom vurdering og motivasjon. [The interplay between assessment and motivation.] In: Dobson S, Eggen A and Smith K (eds) *Vurdering, prinsipper og praksis*. Oslo: Gyldendal, pp.23–39.
- Smith K (2011) Professional development of teachers – A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation* 37(1): 55–61.
- Steffensen K and Ziade SE (2008) *Skoleresultater 2008: En kartlegging av karakterer fra grunnskoler og videregående skoler i Norge*. [School results 2008: A study of grades in Norwegian schools.] Report 2009/23. Kongsvinger: Statistics Norway.
- Svennberg L, Meckbach J and Redelius K (2014) Exploring PE teachers' 'gut feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review* 20(2): 199–214.
- Thorburn M (2008) Articulating a Merleau-Pontian phenomenology of physical education: The quest for active student engagement and authentic assessment in high-stakes examination awards. *European Physical Education Review* 14(4): 263–280.
- Tinning R (2010) *Pedagogy and Human Movement: Theory, Practice, Research*. London: Routledge.
- Troullioud DO, Sarrazin PG, Martinek JJ, et al. (2002) The influence of teacher expectations on student achievement in physical education classes: Pygmalion revisited. *European Journal of Social Psychology* 32(5): 591–607.
- Tveit S (2014) Educational assessment in Norway. *Assessment in Education: Principles, Policy and Practice* 21(2): 221–237.
- Utdanningsdirektoratet (2011) *Grunnlagsdokument: Satsingen Vurdering for læring 2010-2014*. [Background document: Assessment for learning 2010-2014.] Oslo: Utdanningsdirektoratet. Available at: <http://www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Nasjonal%20satsing/2/Grunnlagsdokument%20for%20satsingen%20Vurdering%20for%201%c3%a6ring%20okt%20%2020111.pdf> (accessed 18 July 2013).
- van der Mars H and Harvey S (2010) Teaching and assessing racquet games using 'play practice'. Part 2: Integrating assessment into teaching. *Journal of Physical Education, Recreation and Dance* 81(5): 35–43.
- Vinje E (2008) *Osloundersøkelsen om vurdering i kroppsøving* [The Oslo Study of Assessment in Physical Education]. Nøtterøy: Ped-media AS.



- Young S (2011) A survey of student assessment practice in physical education. *Strategies: A Journal for Physical and Sport Educators* 24(6): 24–26.
- Zhu X (2015) Student perspectives of grading in physical education. *European Physical Education Review*. Epub ahead of print 10 February 2015. DOI: 10.1177/1356336X15569628.

**Author biography**

**Petter E Leirhaug** is a PhD student in Sport Science at the Norwegian School of Sport Sciences, Oslo, Norway. He is also a lecturer and researcher in sport sciences and teacher education at the Sogn og Fjordane University College, Sogndal, Norway.



**Artikkel 3:**

**'The grade alone provides no learning': Investigating assessment literacy among Norwegian physical education teachers**

*Asia-Pacific Journal of Health, Sport and Physical Education* (2016), 7(1), s. 21-36.



## 'The grade alone provides no learning': investigating assessment literacy among Norwegian physical education teachers

Petter E. Leirhaug<sup>a,b</sup>, Ann MacPhail<sup>c</sup> and Claes Annerstedt<sup>d</sup>

<sup>a</sup>Norwegian School of Sport Sciences, Oslo, Norway; <sup>b</sup>Sogn og Fjordane University College, Sogndal, Norway;

<sup>c</sup>Department of Physical Education and Sport Sciences, The University of Limerick, Limerick, Ireland;

<sup>d</sup>Department of Food and Nutrition, and Sport Science, The University of Gothenburg, Gothenburg, Sweden

### ABSTRACT

This paper explores the four inter-dependent elements of assessment literacy proposed by Hay and Penney [(2013). *Assessment in physical education. A sociocultural perspective*. New York: Routledge] – assessment comprehension, assessment application, assessment interpretation and critical engagement with assessment. More specific, the study reported in this paper addresses how Norwegian physical education teachers reflected assessment literacy in descriptions and discussions of their assessment practice. Twenty-three physical education teachers from six upper secondary schools in Norway participated in focus groups. Analysis and discussion are informed by the four elements of assessment literacy. Findings demonstrate a general need to enhance assessment literacy among the teachers, with particular focus on dialogue with students and critical engagement with assessment. Acknowledging assessment literacy as an ongoing process, the study suggests that it may be more effective to consider 'preconditions' than 'elements' of assessment literacy for a physical education teacher to be considered as acting assessment literate.

### KEYWORDS

assessment literacy;  
assessment reform; focus  
group; Norway; physical  
education

### Introduction

Recognising assessment as 'one of the most fraught and troublesome issues physical educators have had to deal with' (López-Pastor, Kirk, Lorente-Catalán, MacPhail, & Macdonald, 2013, p. 57), the aim of this paper is to explore assessment literacy among physical education teachers. The last decade has shown a remarkable growth in studies focusing on assessment in physical education and how assessment practice may affect students' learning. Recent reviews, however, indicate that the area is still weakly conceptualised and the empirical research relatively sparse (Chan, Hay, & Tinning, 2011; Georgakis & Wilson, 2012; López-Pastor et al., 2013). Through analysis of six focus groups with Norwegian physical education teachers, this study contributes to 'the need for empirical work focusing on teachers' existing assessment literacies' (Hay & Penney, 2013, p. 81).

While the term 'literacy' was originally used to refer to the ability to read and write, it has become more common and applies to a wider range of fields, such as 'cultural literacy'

and ‘computer literacy’. In Australia, ‘critical health literacies’ have recently gained attention from researchers in connection with a new curriculum in Health and Physical Education (Corcoran & Renwick, 2014). With inspiration from phenomenological and existential philosophy, ‘physical literacy’ has already entered discussions on physical education (Capel & Whitehead, 2013; Whitehead, 2010).

‘Assessment literacy’ has a longer history in educational literature, and has undertaken varied developments since Stiggins (1991) defined it as understanding the principles of sound assessment, with teachers meeting standards of high-quality classroom assessment. Summing up main developments, Popham (2009) concludes that ‘it seems that assessment literacy is a commodity needed by teachers for their own long-term well-being, and for the educational well-being of their students’ (p. 11).

Hay and Penney (2013) define assessment literacy for teachers as ‘the development of knowledge and capacities to implement assessment and interpret the outcomes of assessment in a manner that is critically aware and that optimises the value of assessment for all students’ (p. 74). This view of assessment literacy reflects the social-constructivist paradigms of knowledge emerging in the twentieth century, and the educational theory developments which place the learner at the centre of the learning process (Klenowski & Wyatt-Smith, 2014; Pryor & Crossouard, 2008).

Norwegian physical education teachers experienced considerable change in their curriculum and assessment regulations with the implementation of a new National Curriculum in 2006 (Arnesen, Nilsen, & Leirhaug, 2013). The reform was followed by officially articulated encouragement to change assessment in schools from an over-emphasis on ‘assessment of learning’ and grading to an emphasis on ‘assessment for learning’ (Utdanningsdirektoratet, 2011). At the same time the reform was underpinned by a new public management-thinking that favoured the accountability assessment perspective (Aasen, 2012). Norwegian physical education teachers have to navigate in an educational landscape which is goal-based with a high degree of local responsibility. The National subject curriculum contains ‘competence aims’ that constitute the basis for all subject assessment, but provides little guidance in the form of regulations or requirements on how to organise and assess students. To a large extent, the physical education teachers have sole responsibility for both providing ongoing assessments that moves student learning forward and assigning the high-stakes grades to students at the end of their secondary level physical education. In practice, this means that the integrity and the quality of the system rely heavily upon the professionalism of teachers in relation to curriculum and assessment standards.

The climate of curriculum and assessment reform is a shared international experience (Berry & Adamson, 2011), and we believe that exploring the Norwegian context will be of interest internationally. This exploration of assessment literacy in physical education is guided by a specific research question: ‘How do Norwegian physical education teachers reflect assessment literacy in descriptions and discussions of their assessment practice?’

The paper is organised in four sections. The first section introduces assessment literacy in physical education and the associated research question that arises. The next section presents the theoretical background and methodological approach, before the third section presents and discusses the findings, followed by the final section that shares conclusions and implications.

## Assessment literacy in physical education

Drawing on Bernstein's (1971) three message systems of education (curriculum, pedagogy and assessment), Hay and Penney (2013) view assessment as a message system. This perspective argues that assessment is a fundamental dimension of the selection, classification, transmission and control of educational knowledge. In different policy and practice contexts, assessment has served very different purposes, but the 'varied purposes of assessment can be broadly categorised within two main endeavours: *accountability* and *learning*' (Hay & Penney, 2013, p. 7; italics in the original). Responding to this, we support Popham (2009) in his distinction between *classroom assessments* and *accountability assessments*. In short, classroom assessments refer to all the procedures that teachers employ to make accurate inferences about what their students know and can do, as well as all formative procedures employed to help student progress. Accountability assessments refer to those measurement devices used by national authorities, school districts or school leaders to ascertain the effectiveness of educational endeavours.

This paper focuses on classroom assessments in physical education, and draws attention to the need for teachers to understand the conditions of assessment efficacy:

... these conditions focus on understanding the purposes of assessment, the contribution of assessment to student learning, the construction of relevant or connected experiences for students, the validity of the interpretations that teachers make and the bases for socially just assessment practices and outcomes. (Hay & Penney, 2013, p. 74)

From this perspective, an investigation of assessment literacy involves asking whether assessment could be considered to be learning oriented, authentic to learning, valid and socially just. Consequently, at least in the Norwegian context, most teachers' engagement in assessment should be in terms of 'assessment for learning' (Georgakis & Wilson, 2012), which was introduced into the assessment discussion by the work of Black and Wiliam (1998). This is not to deny that the assessment literate teacher understands the accountability assessment perspective and acknowledges possible effects within and beyond schools.

Furthermore, Hay and Penney (2013) argue that students need to be assessment literate. Consequently, the awareness of this and how to promote assessment literacy among students becomes part of the assessment literacy for physical education teachers. Noting that not all students are able to read the social cues about how to participate appropriately in assessment practices (Pryor & Crossouard, 2008), it becomes the responsibility of teachers to help students to read, realise and move between the multiple discourses of learning and assessment. The teachers subsequently need to be literate in these discourses.

As Willis, Adie, and Klenowski (2013) reminded us, assessment literacy 'will be understood differently depending on the view of learning embedded in the cultural and policy context and so is closely associated with theories of learning' (pp. 244–245). Drawing on Freebody and Luke (2003), Hay and Penney (2013) argue that 'the capacity to actively participate in a social field is fundamental to what it means to be literate' (p. 71). They proposed four inter-dependent elements of assessment literacy in physical education:

- *assessment comprehension* – focusing on knowledge and understanding of assessment expectations and conditions of efficacy.

- *assessment application* – focusing on the conduct of assessment in terms of either teacher implementation or student engagement.
- *assessment interpretation* – focusing on making sense of and acting on the information that is collected through assessment practices, including traversing and negotiating the social relations of assessment.
- *critical engagement with assessment* – focusing on awareness of the impact or consequences of assessment and challenging the ‘naturalness’ of assessment practices, performances and outcomes. (Hay & Penney, 2013, p. 73; italics in the original)

These elements guided the data analysis of the study reported in this paper, and the findings section elaborates and discusses each of the four elements of assessment literacy in relation to the practices of Norwegian physical education teachers.

### **Research on assessment in physical education with a view to Norway**

Considering the substantial interest in assessment in general education research, Hay (2006) noted that it was ‘strange that assessment in physical education has been somewhat neglected by academic researchers’ (p. 313). Hay and Penney (2013) demonstrate how research into physical education assessment has highlighted ‘the often trivial approaches to assessment by teachers’ (p. 81). Although alternative classroom assessment and the potential positive benefits for teaching and learning in physical education have been advocated for at least two decades in literature, the assessment practices in physical education appear slow to change (Georgakis & Wilson, 2012; Leirhaug & MacPhail, 2015; López-Pastor et al., 2013).

Assessment practices in Norwegian physical education are not well documented (Jonskås, 2010), but it has been established that the 2006 reform made assessment one of the main discussion areas among physical education teachers, mostly because they found the new regulations challenging (Arnesen et al., 2013). In an evaluation report on the development of individual assessment in four school subjects (Norwegian, English, mathematics and physical education), physical education emerges as somewhat different from other subjects. Compared to teachers in the other three subjects, physical education teachers reported that formative assessment to a lesser extent is integrated in their teaching practice, they more seldom set clear goals for student learning and they do not use assessments to revise and adjust their own teaching (Sandvik et al., 2012). Other studies, which have been carried out among Norwegian physical education teachers after the 2006 reform, reveal great differences in assessment practices, which generates problems regarding validity, fairness and comparability (Arnesen et al., 2013; Vinje, 2008). Teachers also report physical education as a difficult subject in which to practice assessment (Prøitz & Borgen, 2010). In Sweden assessment practices were found to be ‘arbitrary’ (Annerstedt & Larsson, 2010, p. 112; Redelius, Fagrell, & Larsson, 2009, p. 257) in regard to alignment with national regulations, and, given the similar curriculum context and practices in Sweden and Norway, this is more than likely also the case for Norway.



## Method

As part of a larger research project about assessment practices in physical education, data for this study were collected across 23 physical education teachers from six upper secondary schools in Norway. The schools were recruited as a convenience sample (Bryman, 2012), with size of school, geographic location and interest in participating in the project as main selection criteria. The main project also included a survey of the students at the schools.

The teachers participated in focus groups at their own school, and in total they (8 females, 15 males, mixed gender in all focus groups) represent all but one of the teachers who taught physical education at the six schools. All teachers had at least one-year specialist training for physical education as part of their teacher education, and ranged from 24 to 61 years of age. Three of the teachers held a Master's degree in physical education.

### Focus groups

Focus groups with the pre-existing group of physical education teachers at each school were chosen because the main project was constructed to allow between-schools comparisons. We also recognised the benefit that teachers 'who know each other may recall common experiences, share half-forgotten memories, or challenge each other on contradictions between what they are *professing* to believe in the group and what they might have said or done outside the group' (Wilkinson, 1998, p. 191; italics in the original). On the other hand, there are challenges when using pre-existing groups appreciating that in 'institutional context (such as workplace or schools), people may be reluctant to express their opinions or discuss their personal experiences in front of colleagues' (Liamputtong, 2011, p. 8). Such challenges did not transpire as the topic was not overly personal, corresponding with the first author's evaluation of the atmosphere of the focus groups and reading of the transcripts.

After a short initial conversation about the status of physical education at the school, asking about equipment and teaching resources, and how teachers' perceived support from the school leadership, the first author presented an open question asking teachers to describe the assessment practices in physical education at their school. A topic guide was developed to map a broad picture of the assessment practices at the schools, and the researcher ensured that the teachers spoke about relevant issues such as understanding of curriculum and assessment regulations, assessment for learning, grading criteria and the future of physical education. The focus groups were structured according to the aims of the main project, and because 'assessment literacy' is an unfamiliar term in Norwegian, they did not include any question or topic referring directly to the term.

All six focus groups were conducted in 2011 and 2012 by the first author in a meeting room or classroom at the teachers' own school and lasted between 45 and 101 minutes. Each focus group was audio-recorded with the teachers' consent and the total recorded time was just over 7 hours. All teachers in the same school took part in the same focus group meaning that some focus groups consisted of two or three teachers. Compared to focus groups with more teachers, the focus groups from the smaller schools could be more accurately termed as group interviews, due to the increased role of the first author in the interactions (Bryman, 2012; Parker & Tritter, 2006).

### **Data analysis**

The audio files from the focus groups were transcribed verbatim and initially subjected to a theoretical thematic analysis (Braun & Clarke, 2006; Bryman, 2012) guided by Hay and Penney's (2013) proposed elements of assessment literacy for physical education teachers (*comprehension, application, interpretation and critical engagement*). After reading the focus group transcripts in order to acquire a sense of the assessment dialogue at each school, the analytical procedure involved three steps: (i) reading and rereading the transcripts to code and separate information into analytical themes informed by the four elements of assessment literacy, (ii) interpreting and comparing the pieces of text in each theme to identify common and conflicting descriptions and discussions among the teachers and across the schools and (iii) a reflective interpretation of the theoretical conceptualisation proposed by Hay and Penney (2013) to see how this could inform the empirical findings and contribute to our understanding of assessment literacy in physical education. Analysis was carried out in Norwegian, and only the extracts of transcripts considered for use in the paper were translated into English.

In conducting the analysis, it proved difficult to code and sort data from the focus groups into only one of the four proposed elements of assessment literacy for physical education teachers. This overlapping structure indicates that the four elements might have even stronger interrelationships than Hay and Penney (2013) suggested. This highlights a need, which we will revisit, for further discussion and development of how to conceptualise and display assessment literacy in physical education. Although we identified themes across the four elements, the next section reports and discusses the findings in accordance with the four elements, with a view to making the discussion of the proposed conceptualisation more visible and ease comparisons with other studies of 'the new and undeveloped idea' (Hay & Penney, 2013, p. 81) of assessment literacy in physical education. The six schools are denoted as A to F, and the teachers at each school are numbered in the order they contributed to the focus group. For example, the third teacher who spoke in the focus group at school B is noted as 'Teacher 3B'.

## **Findings and discussion**

### **Assessment comprehension**

This first element of assessment literacy focuses on the physical education teachers' knowledge and understanding of the official curriculum and regulations covering assessment practice, of assessment to be oriented towards the process of learning (assessment for learning), and the importance of developing assessment literacy among students. Notably, all teachers in the study expressed knowledge of relevant education policy documents and the national curriculum. This was evident through their comments on the content of physical education, examples of competence aims from the physical education curriculum and referencing some of their practices to regulations in the national policy documents.

To illustrate how the teachers understood the assessment regulations, we allow room for a larger excerpt from one focus group where the first author openly asked the teachers to describe their assessment practice:

Teacher 2A: We started out with new rules which came, new regulations on the assessment ... So we have made a type of an overall, or general, plan for [name of upper secondary school]. What we as teachers value high, then.

...

Teacher 2A: I try to follow this plan, on the assessment part, as a starting point. And then we have started to work a little with ...

Teacher 1A: [follow up] ... matrices.

Teacher 2A: Yes, matrices for each ... sport, which is used to assess how the student is doing in the lessons, or how the student performs in various sports.

Researcher: Could you just briefly try to describe how such a matrix ... works?

Teacher 1A: Yes, we have one for every sport. In one column we set up relevant skills for chosen sports and dances, and then we provide descriptions of performance level in three columns; ... skill mastering below average, average and above average.

...

Teacher 1A: The aim is that students should also be able to see this, so they, of course, both recognise where they are and see what it will take to get to the next level. It's kind of a step in the direction of making the assessment somewhat concrete, so we can point to it and that students also can see it.

While the excerpt demonstrates how the teachers have started to implement the new national guidelines for assessment, it also illustrates how assessment is decentralised to each school to determine how to meet the requirements. At this school teachers had chosen to develop matrices to clarify assessment criteria for the students. Matrices are neither mentioned nor suggested in the national regulations, conveying that the assessment reform is being enacted as local practices.

Comparisons across the six schools confirm clear differences in the ways physical education teachers understand the criterion-referenced assessment system, the purposes of assessment and how they interpret the regulations. An example was whether or not to consider 'effort' in what is valued in assessment in physical education, a topic discussed in all focus groups. Referring to the fact that the 2006 reform stated that effort should not be included in subject competence assessments, teacher 1B explained why they emphasise skills and knowledge at the expense of effort: It 'is much about being loyal in relation to assessment regulations and curriculum'. In contrast teacher 3E stated:

You cannot play football without putting in some sort of *effort*. I do not quite understand, it [effort] has always been there and will always be there. I think it was just crazy, then, when it was underlined that we should not assess students' effort. It's kind of at the core of our subject ...

Significant between-school differences also appear with regard to the importance of assessment being oriented towards enhancing student learning. While teachers in two schools referred to 'assessment for learning' to explain what they were practicing with respect to assessment in physical education, the same concept appeared to be unfamiliar to some of the teachers at the other schools when the researcher asked directly about it:

Researcher: If I say assessment for learning in physical education, what does that make you think about?

Teacher 2A: It must be towards something ... you mean *for* learning? Well, the first thought that comes to my mind is motivation. Directed towards something, I am going to improve someone. It is motivation towards something, isn't it?

The 2006 reform was followed by minimal support for implementation at the teacher level (Aasen, 2012; Arnesen et al., 2013), which may somewhat explain the observation from this study that the national guidelines and competence aims in the physical education curriculum appeared to be interpreted into local, rather than national, practices. Traditions and assessment culture at the schools, as well as the individual teachers' beliefs and values, appeared to have influenced how the teachers understood and tried to cope with what they had perceived as changes in assessment and documentation requirements.

A colleague of teacher 2A developed the conversation and illustrated how differences in understandings not only appear between schools but within schools:

Teacher 1A: No, assessment for learning? ... then it is about this formative assessment ['*underveisvurdering*' in Norwegian] then; a bit more detailed and informative assessment. The grade alone provides no learning; learning stems from the elaboration and the feedback that we provide along the way, ... that is what assessment for learning must be.

Incidents where colleagues help, moderate or negotiate each other's understanding are evident in all six focus groups. Such incidents indicate how important it is to allocate time for teachers to come together regularly and share pedagogical thinking and experiences.

In regard to the importance of developing assessment literacy among students, several teachers alluded to the value of engaging students in their own learning and some gave examples of student self-assessment and how they, as teachers, try to be open and clear in communication around assessment practices. However, none of the teachers provided arguments based on the importance of empowering the students with understanding of assessment processes deeper than clarifying learning goals, sharing criteria for success and making sure that students are well informed about the assessment process and what they can do to improve.

Acknowledging the importance of self-assessment when prompted, teacher 1D admitted: 'I do ask the students too little about what they think about themselves.' Several teachers provided examples of how they recently had begun to explore self-assessment:

Teacher 1B: I would say, we have been in a bit of an exploring phase, so we have tried a bit differently. One example can be: ... thumbs up – if you feel that you understand and can do, thumbs to side – if you feel you are like in the

middle but 'on the way', and thumbs down – if you think you are not doing so good.

The example provided by teacher 1B represents an exception. Most teachers attached self-assessment to grading and being used as a means to 'regulate reality orientation' (Teacher 2F) of the student's awareness of his or her level. Describing self-assessment the teachers often referred to one particular competence aim in the physical education curriculum stating that a student should be able to plan, carry out and assess a period of self-guided exercise. The physical education teachers appeared reluctant to engage with student peer-assessment. Prompted on the issue, teacher 2A said straight forward 'No, I have never done that.' Only one out of the 23 teachers provided a real positive example of enacting peer-assessment. Given that peer-assisted learning is one of the more thoroughly researched areas in physical education, reporting mostly positive student learning experiences (Ward & Lee, 2005), the potential for developing peer-assessment is considerable.

### **Assessment application**

This element of assessment literacy foregrounds the capacity and ability of teachers to implement and carry out efficacious assessment practices. Drawing attention to the long excerpt above from the focus group where the physical education teachers presented the matrices they were working on, the capacity of this element becomes highlighted when they were prompted to say something about the actual use:

Researcher: Have you tried to use these [matrices] with the students?

Teacher 1A: I have not tried to use them in collaboration with students yet, ... I have not.

The application of the matrices has not yet been realised. Later in the same focus group, a colleague expresses that he is afraid that 'they can end up making a higher workload for themselves' (Teacher 2A), but is eager to add that he thinks the idea with the matrices is good.

In discussing what teachers actually do to assess, the teachers in five of the six schools conduct pen-and-paper tests in physical education. Three schools have a system for conducting prepared conversation with students about once a year. All six schools use 'Fronter' (an ICT learning platform) to provide information about the subject, but only two schools use it for documenting feedback and grading. In regard to *assessment application*, most of the time in the focus groups focused on describing and discussing the use of practical techniques or physical tests. Two schools use a number of skill-tests in selected sports, two schools conduct more traditional 'physical fitness tests', one school utilises both skill-tests and 'physical fitness tests', while the final school has ceased using tests.

Some of the teachers pointed to the importance of more formalised assessment situations and that students should know when they are being assessed. However, teachers in all six focus groups made reference to the importance of teacher ongoing observations of student performances during the lessons, and, in terms of grading student competence,

they were willing to negotiate the individual student's results in planned assessment situations with their ongoing observations. They referred to this as more comprehensive assessment or assessment that was conducted throughout a block of lessons.

### ***Assessment interpretation***

A number of teachers expressed the view that physical education is a subject with particular difficulties or challenges when it comes to making judgements in assessment:

Teacher 2D: Generally, I think it is difficult to do assessment in physical education. How good are you when you are good? What level is the right one to get the grade 3, how good should one then be in the different sports?

Such expressions were not contested in the focus groups with a complex set of reasons for this perception arising. Teacher 4F shared that physical education 'is the subject of assessing moments, or "short-cuts", right? Twenty seconds, I observed Lisa who did some skilled and entertaining moves'. The challenge of assessing short-cuts connects to what teacher 1D suggested when he stated that it can be 'unbelievably hard' to look for the deeper learning of each student, when you have 'more than 20 students supposed to move around'. In addition, the teachers seemed to share the view expressed by teacher 4C that the 'curriculum is very ambitious in a subject which only has 90 minutes a week', and allude to the belief that the competence aims, which forms the base of subject assessment, are broad and vague. Interestingly, at school C, assessment was part of the explanation when the teachers reflect on why they do not teach the curriculum expected amount of outdoor life/friluftsliv:

Teacher 3C: ... it also has something to do with assessment. I do think it's not so easy to assess skills in friluftsliv [outdoor life]. It is all more and more about tutorials and tips, it is handy with a little practice in outdoor activities but difficult to do assessment in.

Somewhat contradictory to the expressed difficulties of assessing students in physical education, teachers tended to be confident that most students were assigned correct grades. However, as already indicated above, the basis for the grading varies, both between and within schools. At four of the schools teachers explicitly believed that the students were satisfied with grading on the basis that teachers do not need to discuss grades with the students.

While much time in the focus groups was spent on discussing grading and how to ensure students were assigned the grade they deserved, none of the teachers used words like 'validity' and 'reliability'. For example, when addressing fairness, the general impression was that teachers were more worried about comparability to other schools, than the validity and possible equity problems with their own judgements in student assessment. This was noted by Prøitz and Borgen (2010) as the main concern of physical education teachers' expression of fairness, with comparability more likely to be regarded as a continuous challenge in a decentralised assessment system where teachers are responsible for assigning high-stake grades.

It has been suggested that summative assessment by teachers will only be accountable and truly comparable if assessments can be shown to be aligned with shared standards of the same curriculum. Klenowski and Wyatt-Smith (2014) suggest that moderation is one way to meet this need, defining moderation as ‘a process of involving teachers in discussion and debate about their interpretations of quality of assessed work’ (p. 74). The physical education teachers in this study appeared to have a high level of individual autonomy but no directive, as well as few opportunities, to connect with colleagues to discuss issues concerning assessment, physical education and pedagogy. The findings also demonstrated the need for teacher learning about validity as a judgment about the degree to which the assessment results are appropriate for making certain educational inference and decisions (Messick, 1989).

*Assessment interpretation* is also about how teachers utilise information sought from assessment ‘so that appropriate curricular and pedagogical adjustment can be made’ (Hay & Penney, 2013, p. 76). Conveying a view of providing feedback continually during practice in physical education lessons, all focus groups provided indirect examples of how the teachers made more or less unconscious pedagogical adjustments based on assessment of where the student is in the learning process. A number of teachers expressed directly how ‘all this feedback is part of the ongoing assessment in the subject, and as such, it is feedback in relation to the overall competence aims’ (Teacher 6F). This indicates an alignment of curriculum and teaching practice that teachers present as somewhat domain specific, supported by examples of how such adjustments can be more sophisticated:

Teacher 2F: I also think that one of those things that is special with physical education, is that it is about movement, it is about skills, practical tasks and learning. It is not necessary to put so much into words of what you do. You can arrange a situation that makes them [the students] in a way progress.

However, none of the teachers described student assessment as something they consciously and more systematically use to inform and adjust their own teaching and assessment practice. As already alluded to, most assessment interpretations discussed among the teachers were attached to questions of accountability, documentation and grading, and can be referred to as traditional assessment (López-Pastor et al., 2013).

### ***Critical engagement with assessment***

In response to this element of assessment literacy, the focus groups did not produce one clear example of a teacher considering power relations in the process of assessment. Except from a few episodic reflections from individual teachers the analysis revealed little, and no systematic or common, self-analysis of their own role in assessment or how traditions, values and beliefs unavoidable influence assessment interpretations and judgements. However, there were many descriptions of episodes which represented possible opportunities to undertake critical reflection upon enacted practice and outcomes. One such example was when a teacher reported the experience of sharing criteria with students resulted in students ‘who really disagree and who think it should be enough to try as good as you can and to show good effort’ (Teacher 1B). The opportunity for the teacher to consider the students’ responses to the feedback and re-examine the assessment

criteria was not pursued by teachers' discussion. In a similar vein, when teacher 3C expressed that 'some students don't like the physical fitness tests', he does not step back and consider if it is necessary for all students to carry out the tests, even if there is no requirement in the assessment regulations that underpin the use of such tests.

### **Assessment literacy in physical education revisited**

Assessment literacy and illiteracy are not only about competence but also about the capacity to 'become' assessment literate in particular settings or situations (Willis et al., 2013). Situation and contexts determine appropriate use of assessment and consequently, the same teacher can be assessment literate in one situation and assessment illiterate in another.

Adopting Hay and Penney's (2013) four inter-dependent elements of assessment literacy allows us to begin interrogating the empirical data in ways, and identify themes, that the data alone would perhaps not suggest. The framework enabled findings to emerge from interpretation, and consequently the proposed framework has proven beneficial with the missing reflection on power relations and teachers' critical engagement with assessment as most evident examples. However, due to the inter-dependent nature of the elements we experienced some challenges in sorting data to only one of the four stipulated elements. Acknowledging assessment literacy as an ongoing process, we suggest that it might be better to propose 'preconditions', rather than 'elements', for a physical education teacher to enact assessment in ways that are to be considered assessment literate.

Expected preconditions for assessment literacy are related to teachers' critical view on their own assessment practice and how it can affect students, their preparation and promotion of students' assessment literacy (which include engaging students in assessing their own and their peers' learning) and sharing interpretations and beliefs with other colleagues from time-to-time to moderate and update their assessment practices. The findings of this study demonstrated that the Norwegian physical education teachers did not fulfil such preconditions. They did not appear to consider student involvement in the process of formulating assessment criteria or include students on discussions of how best to assess in physical education. In at least four of the schools, the teachers continued to rely on the teaching of sport skills and the associated traditional end-of-unit-assessment, which probably 'can also be interpreted as an alibi for the "gut-feeling" when justifying grades' (Svennberg, Meckbach, & Redelius, 2014, p. 209). This observation of persisting traditional approaches to assessment is supported by recent findings in two Masters' theses that focused on students' views on assessment in Norwegian physical education (Græsholt, 2011; Palm, 2014).

### **Conclusion and implications**

In this paper, we have explored assessment literacy in physical education through an investigation of focus groups with Norwegian physical education teachers. The analysis revealed assessment as an established discourse, or maybe even what we can call a heated discourse, among the physical education teachers. All teachers in the study demonstrated knowledge of the national curriculum and conveyed a level of transition in



engaging with assessment. Aware of new regulations, and to different degrees inspired and challenged by the sharpened focus of assessment and accountability (Aasen, 2012), the physical education teachers were focused on clarifying the assessment criteria and finding ways to document feedback to students. In 'contrast to traditional assessment practices in physical education that are more often than not an afterthought to the curriculum and pedagogy process and/or remain essentially disconnected from what is fundamentally taught or intended to be learnt' (Hay & Penney, 2013, p. 75), several teachers explored the use and possible benefits of alternative assessment practices (López-Pastor et al., 2013). However, the teachers did not appear to approach this in a systematic and planned way, but rather deferred to the more reactive response associated with traditional assessment. Acknowledging only a few examples of engagement of students in assessing their own and their peers' learning, as well as a lack of critical awareness and reflexive considerations of power relations, different opportunities, and possible negative outcomes of enacted assessments, we conclude that the physical education teachers' descriptions and discussions of assessment practices do not reflect central preconditions of assessment literacy.

The findings in this study have raised concerns about validity and comparability of both summative and formative assessment in Norwegian physical education with possible unjust and negative consequences for students. It is also worrisome that everyday life in schools appears to create few opportunities for physical education teachers to connect with colleagues to discuss issues concerning assessment. To ensure validity and comparability in a decentralised school system like that in Norway, it is essential that teachers are well-educated in assessment and acknowledge assessment literacy as an ongoing process in which preconditions for quality assessment are negotiated with curriculum and students.

Further, the study highlights a need to focus on assessment literacy in pre-service, as well as in-service, education of physical education teachers. We emphasise three issues in particular for development. First, we reaffirm the Georgakis and Wilson's (2012) observation of the lack of practical examples and pedagogical models for formative assessment in physical education, and how formative assessments contribute to summative assessments such as grading. It is of the utmost importance to develop subject-relevant material. Second, it is necessary for teachers to work on creating awareness of the importance of developing assessment literacy for students. It is not possible for teachers to make assessment transparent and understandable to students without enhancing assessment literacy among their students. Third, the perspective included in Hay and Penney (2013) element of 'critical engagement with assessment' requires particular attention in physical education teachers' professional learning.

The study being conducted solely in a Norwegian context is, of course, a limitation. However, compared to reports of physical education teachers' assessment practices in many other countries (Georgakis & Wilson, 2012; López-Pastor et al., 2013; Redelius & Hay, 2009), there are reasons to believe that the Norwegian context does not differ extensively. Another limitation is that the study only focuses on teachers' assessment literacies, and there is a need also to investigate preconditions of students' assessment literacies.

Recognising assessment literacy as an ongoing process constantly negotiating the contexts, assessment literacy in physical education can only be enhanced through the daily practice of physical education teachers, most likely with some level of support from the national level with, for instance, examples of assessment practices. With a heightened

focus on assessment in teacher education, Popham (2009) noted that ‘in a decade or two, the assessment literacy of the nation’s teaching force is bound to be substantially stronger’ (p. 5). However, this may be a vain hope since physical education teacher education in Norway, like in many other countries, is ‘neither shaking nor stirring’ the students (Moen & Green, 2014). Moen and Green (2014) refer to the fact that the understanding and development of physical education teachers’ professional identities is anchored in a sports and performance culture, rather than in critical reflection and educational theory focusing on student learning (Dowling, 2011). This view is also supported by Backman and Larsson (2013) who, in their study of physical education teacher education, point to assessment as one of the areas that needs to be developed. Our study supports the need for such development and suggests that assessment literacy provides the framework and scope in which such work can reside.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Notes on contributors

**Petter E. Leirhaug** is a PhD student in Sport Science at the Norwegian School of Sport Sciences, Oslo, Norway. He is also a Lecturer and Researcher in sport sciences and teacher education at the Sogn og Fjordane University College, Sogndal, Norway. His research interest and work is focussed on assessment, learning and curriculum development in physical education, teacher education and outdoor education/‘Nordic friluftsliv’.

**Ann MacPhail** is Associate Professor and Head of the Department of Physical Education and Sport Sciences at the University of Limerick, Ireland. Her teaching and research interests revolve around teacher education, young people in sport, assessment and curriculum development in physical education. Ann is Co-Director of the Physical Education, Physical activity and Youth Sport Research Centre (PE-PAYS). She is Associate Editor for the Physical Education and Sport Pedagogy and a member of the Editorial Board for the Journal of Teaching in Physical Education.

**Claes Annerstedt** is Head of the Department of Food and Nutrition and Sport Science at the University of Gothenburg, Sweden. He is a former Professor of pedagogy and physical education from Norwegian School of Sport Sciences in Oslo, Norway, and his main teaching and research interests are curriculum development, teaching and learning in physical education, as well as in sport coaching and teacher education.

### References

- Aasen, P. (2012). Accountability under ambiguity. Dilemmas and contradictions in education. In A. L. Østern, K. Smith, T. Ryghaug, T. Krüger, & M. B. Postholm (Eds.), *Teacher education research between national identity and global trends* (pp. 77–92). Trondheim: Akademika.
- Annerstedt, C., & Larsson, S. (2010). ‘I have my own picture of what the demands are ...’ Grading in Swedish PEH – Problems of validity, comparability and fairness. *European Physical Education Review*, 16(2), 97–115.
- Arnesen, T. E., Nilsen, A. K., & Leirhaug, P. E. (2013). ‘Den læreplanen som ikkje kan tilpassast mi undervisning, finst ikkje.’ Vurdering og undervisning i kroppsøving etter Kunnskapsløftet [Assessment and teaching in physical education after the reform the Knowledge Promotion]. *Tidsskriftet FOU i praksis*, 7(3), 9–32.

- Backman, E., & Larsson, L. (Eds.). (2013). *I takt med tiden? Perspektiv på idrottsl r rutbildningen i Skandinavien* [Fit for the time? Perspectives on the physical education teacher education in Scandinavia]. Lund: Studentlitteratur.
- Bernstein, B. (1971). *Class, codes and control*. London: Routledge.
- Berry, R., & Adamson, B. (Eds.). (2011). *Assessment reform in education: Policy and practice*. Dordrecht: Springer.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–73.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press.
- Capel, S., & Whitehead, M. (Eds.). (2013). *Debates in physical education*. London: Routledge.
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in Physical Education. *Asia-Pacific Journal of Health, Sport and Physical Education*, 2(1), 3–19. doi:10.1080/18377122.2011.9730340
- Corcoran, T., & Renwick, K. (2014). Critical health literacies? Introduction. *Asia-Pacific Journal of Health, Sport and Physical Education*, 5, 197–199. doi:10.1080/18377122.2014.940807
- Dowling, F. (2011). ‘Are PE teacher identities fit for postmodern schools or are they clinging to modernist notions of professionalism?’ A case study of Norwegian PE teacher students’ emerging professional identities. *Sport, Education and Society*, 16(2), 201–222.
- Freebody, P., & Luke, A. (2003). Literacy as engaging with new forms of life: The ‘four roles’ models. In G. Bull & M. Anstey (Eds.), *The literacy lexicon* (2nd ed., pp. 51–66). Frenchs Forest, Australia: Prentice Hall, Pearson Education.
- Georgakis, S., & Wilson, R. (2012). Australian physical education and school sport: An exploration into contemporary assessment. *Asian Journal of Exercise & Sport Science*, 9(1), 37–52.
- Gr sholt, S. A. (2011). *Elevvurdering og vurderingskultur i kropps ving* [Assessment culture in physical education] (Unpublished master’s thesis). The Norwegian School of Sport Sciences, Norway.
- Hay, P. (2006). Assessment for learning in physical education. In D. Kirk, D. Macdonald, & M. O’Sullivan (Eds.), *The handbook of physical education* (pp. 312–325). London: Sage.
- Hay, P., & Penney, D. (2013). *Assessment in physical education. A sociocultural perspective*. New York: Routledge.
- Jonsk s, K. (2010). *En kunnskapsoversikt over FOU-arbeid innen kropps vingsfaget i Norge fra januar 1978 – desember 2010* [An overview of research in physical education in Norway from January 1978 to December 2010]. Oslo: Norges Idrettsh gskole.
- Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. London: Sage.
- Leirhaug, P. E., & MacPhail, A. (2015). ‘It’s the other assessment that is the key’: Three Norwegian physical education teachers’ engagement (or not) with assessment for learning. *Sport, Education and Society*, 20(5), 624–640.
- Liamputtong, P. (2011). *Focus group methodology. Principles and practice*. London: Sage.
- L pez-Pastor, V. M., Kirk, D., Lorente-Catal n, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18, 57–76.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- Moen, K. M., & Green, K. (2014). Neither shaking nor stirring: A case-study of reflexivity in Norwegian physical education teacher education. *Sport, Education and Society*, 19(4), 415–434.
- Palm, T. (2014). *Assessing with Foucault* (Unpublished master’s thesis). The Norwegian School of Sport Sciences, Norway.
- Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research and Method in Education*, 29(1), 23–37.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48, 4–11.

- Prøitz, T. S., & Borgen, J. S. (2010). *Rettferdig standpunktvurdering – det (u)muliges kunst? Læreres setting av standpunktkarakter i fem fag i grunnopplæringen* [Fair grading – An (im)possible art?]. Report no. 16/2010. Oslo: NIFU STEP.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1): 1–20.
- Redelius, K., Fagrell, B., & Larsson, H. (2009). Symbolic capital in physical education: To be, to do or to know? That is the gendered question. *Sport, Education and Society*, 14(2), 245–260.
- Redelius, K., & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15(3), 275–294.
- Sandvik L. V., Engvik G., Fjørtoft, H., Langseth, I.D., Aaslid, B.E., Mordal, S., & Buland, T. (2012). *Vurdering i skolen. Intensjoner og forståelser* [Assessment in school. Intentions and understanding]. Research report. Trondheim: NTNU, SINTEF. Retrieved December 14, 2012, from [www.udir.no/Upload/Rapporter/2012/fivis.pdf?epslanguage=no](http://www.udir.no/Upload/Rapporter/2012/fivis.pdf?epslanguage=no)
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534–539.
- Svennberg L., Meckbach J., & Redelius, K. (2014). Exploring PE teachers' 'gut-feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20(2), 199–214.
- Utdanningsdirektoratet. (2011). *Grunnlagsdokument: Satsingen Vurdering for læring 2010–2014* [Background document: Assessment for learning 2010–2014]. Oslo: Utdanningsdirektoratet. Retrieved June 18, 2013, from [www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Nasjonal%20satsing/2/Grunnlagsdokument%20for%20satsingen%20Vurdering%20for%201%3%a6ring%20okt%20%2020111.pdf](http://www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Nasjonal%20satsing/2/Grunnlagsdokument%20for%20satsingen%20Vurdering%20for%201%3%a6ring%20okt%20%2020111.pdf)
- Vinje, E. (2008). *Oslundersøkelsen om vurdering i kroppsøving*. Nøtterøy: Ped-media AS.
- Ward, P., & Lee, M. A. (2005). Peer-assisted learning in physical education: A review of theory and research. *Journal of Teaching in Physical Education*, 24(3), 205–225.
- Whitehead, M. (2010). *Physical literacy throughout the life course*. London: Routledge.
- Wilkinson, S. (1998). Focus group methodology: A review. *International Journal of Social Research Methodology*, 1(3), 181–203.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256. doi:10.1007/s13384-013-0089-9

**Artikkel 4:**

**'It's the other assessment that is the key': three Norwegian physical education teachers' engagement (or not) with assessment for learning**

*Sport, Education and Society* (2015), 20(5), s. 624-640.



# **‘It’s the other assessment that is the key’: three Norwegian physical education teachers’ engagement (or not) with assessment for learning**

Petter E. Leirhaug<sup>a\*</sup> and Ann MacPhail<sup>b</sup>

<sup>a</sup>*Norwegian School of Sport Sciences, Oslo, Norway;* <sup>b</sup>*Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland*

The international agenda for assessment continues to convey a growing interest in assessment for learning (AfL) as a tool to support learning and enhance teaching. Complementing this, the recent literature on assessment in physical education acknowledges the need for physical educators to integrate AfL into their teaching and assessment practice as an important part of the future development of the subject. Appreciating that physical education must be recognized as part of the larger movement culture in society and is a place to learn about movement culture, this study explores how AfL is understood and enacted by physical education teachers and the extent to which such enactment complements or challenges learning movement cultures within physical education. This study shares how three Norwegian physical education teachers used AfL to term what they were practicing with respect to assessment in physical education. We follow the interactions of the selected teachers throughout focus groups, using the empirical data as our ‘dialogue partner’ in reconstructing and discussing their assessment stories. We conclude that the need of embedding AfL in learning theory may well be one of the strongest challenges to enacting AfL in physical education. We acknowledge that not only are most existing theories of learning defined cognitively, but also that learning connected to physical education and activity is, to a large extent, practical and embodied, and also linked to the powerful discourses of sport and related areas such as health.

**Keywords:** *Assessment for learning; Formative assessment; Physical education; Teacher stories; Norway*

## **Introduction**

Assessment for Learning (AfL), in contrast to the more traditional ‘assessment of learning’, has more recently gained considerable interest in research and development on assessment in physical education. AfL is viewed as fundamental to enhancing student learning and developing physical educators’ teaching and assessment practice, and as such is deemed to be an important element of physical education’s future (Chan, Hay, & Tinning, 2011; Georgakis & Wilson, 2012; Hay &

---

\*Corresponding author. Norwegian School of Sport Sciences, Oslo, Norway. Email: [petter.erik.leirhaug@nih.no](mailto:petter.erik.leirhaug@nih.no)

Penney, 2013; López-Pastor, Kirk, Lorente-Catalán, MacPhail, & Macdonald, 2013; MacPhail & Halbert, 2010; Ni Chróinín & Cosgrave, 2013). That said, assessment in physical education has been an area with little consensus, questionable validity and a slow ability to change (Annerstedt & Larsson, 2010; Arnesen, Nilsen, & Leirhaug, 2013; Hay & Penney, 2013). The historical legacy of physical education being closely linked to the practice of sport and performance culture does not necessarily correspond with the advances in educational theories of knowledge and learning (Kirk, 2010; Quennerstedt, Öhman, & Armour, 2014), and the need of embedding 'alternative assessment' in a learning culture (Black & Wiliam, 2009; Shepard, 2000).

The focus in school physical education is argued to be 'more on activity than on learning' (Annerstedt, 2008, p. 316), often linked to a taken-for-granted argument of how the subject contributes to a healthy and lifelong interest in sport and physical activity. Among others, Green (2014) concludes that there is little research to support the notion that physical education contributes to such a lifelong interest in sport and physical activity. Interestingly, Säfvenbom, Haugen, and Bulie (2014) view the mandatory aspects of physical education as a double-edged sword which also can 'prevent the adolescents from leaving a harmful context and thus produce learned helplessness and an overall negative attitude to movement contexts' (p. 6). Similar to many other countries, physical education in Norway 'seems to favour those who are already involved in movement activity and those who are involved in competitive youth sports in particular' (Säfvenbom et al., 2014, p. 15). It was partly to move away from such a 'meritocratic mode' that Bart Crum (1993), more than 20 years ago, introduced 'movement culture' as the core of physical education. Rather than making 'unrealistic effect claims' (Crum, 1993, p. 352) on behalf of physical education, Ward (2014) agrees that physical education must be recognized as part of the larger movement culture in society, and a place to learn about this, 'Physical Education should be a learning context in which pupils develop a personal movement identity which inherently involves utilising a range of Sporting activities to enable them to become critical and lifelong consumers of Movement Culture' (p. 577).

Calling for lifelong and life-wide learning in physical education, Hay and Penney (2013) recognize AfL:

as a tool that can be utilised to generate discussions and new thinking about learning and learning opportunities in physical education within and beyond schools. It is concerned with quality and equitable learning opportunities and experiences now and in the future, for all students. (p. 110)

In a recent review on 'alternative assessment' in physical education, López-Pastor et al. (2013) observed 'that such innovative practice is, however, far from regular, integral, widespread and educationally productive' (p. 73). Drawing attention directly to AfL, Georgakis and Wilson (2012) highlighted two main challenges for teachers wishing to engage with AfL; (1) a lack of practical examples and (2) a tendency in literature to simply reiterate to the reader that assessment in physical



education needed to be 'authentic'. In interviewing 17 teachers of physical education and school sport in Australia, Georgakis and Wilson (2012) reported that physical education and school sport teachers did not use the 'assessment for learning' term or discuss links between assessment and learning.

With this as a backdrop, the study in this paper shares instances where Norwegian physical education teachers referred to 'assessment for learning' (unprompted) as what they were doing, or attempting to do, in their engagement with assessment practice in school physical education. Taking this empirical finding as our starting point, the purpose of the study is to explore AfL in physical education through Norwegian physical education teachers' narratives. The main research question that directs this study is 'How is AfL motivated, understood and enacted by physical education teachers who are trying to integrate AfL into their teaching and assessment practice?'

Before sharing the assessment stories of three Norwegian physical education teachers, we map the development of AfL and research on AfL in physical education. We then present the Norwegian study context and discuss the methodological approach.

### **Assessment for learning**

'Formative assessment' has undergone considerable development since Scriven (1967) coined the term related to curriculum and school reforms. It was not before the late 1980s that 'the idea that classroom assessment practices could both afford and constrain student learning began to gain widespread acceptance' (Wiliam, 2011, p. 13), and during the 1990s that 'assessment for learning' became the preferred phrase to 'formative assessment'. The purpose of supporting student learning and learning opportunities defines AfL, allowing AfL to retain both summative and formative assessments. The validity and reliability questions are still relevant to examining the extent to which AfL practices support student learning and achievement (Gardner, 2012). Well-planned AfL informs the most effective, meaningful and worthwhile instruction strategies to improve teaching and the subsequent student learning experiences (Tannehill, van der Mars, & MacPhail, 2013).

Ongoing discourse surrounds how to theorize and understand AfL, and how to realize its considerable promises (Bennett, 2011; Gardner, 2012; Wiliam, 2011). We favour the view that AfL is embedded in a social-constructivist view of learning which places the learner at the centre of the learning process (Shepard, 2000; Hay, 2006). We subsequently adopt the much cited definition of AfL as, 'the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there' (Assessment Reform Group, 2002, p. 2).

In terms of classroom strategies and practical techniques that teachers can use to improve the quality of instructional decisions, the AfL literature agrees on some key

Table 1. Five key strategies of AfL, adapted from Wiliam and Thompson (2008)

	Where the learner is going?	Where the learner is right now?	How to get there?
Teacher	(1) Clarifying learning intentions and criteria for success	(2) Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	(3) Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	(4) Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	(5) Activating students as the owner of their own learning	

principles. Table 1 presents the five key strategies, numbered one to five, we use as a base for analysis.

Drawing on Shulman's (1986) distinction between subject knowledge and pedagogical content knowledge, Bennett (2011) argues that quality formative assessment depends on deep subject domain understanding and highlights two implications. First, the importance of interpreting assessment information meaningfully and being able to arrange suitable feedback. Second, and partly as a consequence of the first implication, 'the intellectual tools and instrumentation we give to teachers may differ significantly from one domain to the next because they ought to be specifically tuned for the domain in question' (pp. 15–16).

The limited number of AfL studies related to physical education reports positive learning movement experiences among students, and indicate that the implementation of AfL can be productive in teacher development and result in improved student learning. On completion of a physical education assessment-planning AfL framework to a number of schools in Ireland, MacPhail and Halbert (2010) reported that both the teachers and students believed that the AfL methodology had improved the quality of student learning. In implementing formative assessment in primary schools, Ni Chróinín and Cosgrave (2013) stated that AfL supported worthwhile experiences in primary physical education, recording a dramatic change:

Teachers' perspectives on assessment changed dramatically as the value of assessment in enhancing the learning process was recognised. As a result of using assessment strategies, the teachers believed that the children learned more in their physical education classes and that they themselves became better teachers of physical education. (p. 230)

While AfL has resulted in positive and welcome changes in student learning within school physical education, this has not been without challenges. Teachers have reported finding the implementation time consuming (MacPhail & Halbert, 2010). It has also been reported that students believe assessment to be boring, subsequently making it difficult for the physical education teacher to engage them 'authentically' in

the process (James, Griffin, & Dodds, 2009). Further, student views of expected outcomes of physical education, and the knowledge produced through participation in physical education (Quennerstedt, 2013), can be quite different from learning outcomes stated in the official curriculum. The lack of alignment and the connected problems for validity and equality in assessments, most often with a reference to Bernstein's (1975) three message systems (curriculum, pedagogy and assessment), is well documented in physical education (Annerstedt & Larsson, 2010; Hay & Macdonald, 2008; Redelius, Fagrell, & Larsson, 2009; Redelius & Hay, 2009). Conscious that the question of validity and the AfL strategies employed are dependent of both the actual curriculum, and the wider political, social and cultural contexts, the next section provides a brief presentation of educational assessment and physical education in Norway in a bid to set the context of the study.

### **The Norwegian context**

In 2006 the Norwegian school reform 'The Knowledge Promotion' introduced new national curriculums for all subjects, leading to an increased focus on assessment in part due to accountability perspectives and neoliberal ideas guiding educational reforms in many capitalist welfare states (Møller & Skedsmo, 2013). At the same time, the new assessment regulations were informed directly by the research of Black and Wiliam (1998), and there was an articulated goal to change assessment in schools from an over-emphasize on assessment of learning to an emphasize of AfL (Utdanningsdirektoratet, 2011). The 2006 reform resulted in student assessment becoming one of the main discussion areas among physical education teachers due to the challenging nature of reconsidering their practices (Arnesen et al., 2013).

The Norwegian physical education curriculum is characterized by a broad content area that can be described as a multi-activity model where the students are to learn about body and health through the use of their bodies. The experiences during physical education classes should prepare young students to know, and be able to participate in, the most significant part of Norwegian movement culture, and motivate lifelong interest in physical activity. The national curriculum's 'competence aims' describe the expected learning outcomes and constitute the basis for all assessment (summative and formative) in the subject.

Students obtain grades from year eight onwards (aged 12 years), and the grading scale is divided into six steps expressed from 1 to 6 (1 expresses very low competence and 6 expresses eminent competence in the subject). The consideration of educational assessment and how to organize teaching is highly decentralized in Norway, with schools expected to develop their own local work plan which includes the criteria for assessment and grading. To a large extent, student assessment is dependent on the assessment of the individual teacher.

## **Methodology**

The authors share the assessment stories of three physical education teachers, 'Geir', 'Ronald' and 'Leif' (pseudonyms). The selection of these teachers was as a result of analyses of focus groups conducted as part of a larger project which aims to study both students' and teachers' perceptions of assessment in physical education across six upper secondary schools in Norway.

The main project included six focus groups with a total of 23 physical education teachers. In two focus groups, three teachers started to talk unprompted about AfL. A decision was made to focus more closely on their interactions to explore their use of the term AfL to describe what they were doing in physical education.

Ronald and Leif worked in the same school and participated in a focus group with four other physical education teachers, while Geir and one other colleague constituted another focus group. Focus groups were facilitated by the first author, recorded with a digital voice recorder and transcribed verbatim. Due to the more active role of the researcher in the focus group Geir contributed to the focus group could be more appropriately termed a group interview (Parker & Tritter, 2006). The focus group (1 hr 7 min) including Geir was conducted in a meeting room at his school while a classroom was used at Ronald's and Leif's school for a focus group (1 hr 40 min).

The first author translated assessment-relevant sections of the transcripts into English. Informed by reflexive interpretation (Alvesson & Sköldbberg, 2009), we have downplayed the importance of collected data as focus group material, and use transcripts as 'dialogue partners' in the following analytical approach. Both authors individually completed a thematic analysis in relation to the research question and AfL key strategies, before sharing their emerging understandings with each other.

Similar to Torrance and Pryor's (1998) investigation of formative assessment through interviews with teachers, we experienced that the individual teacher can represent ambiguous and sometimes contradictory views of assessment, as well as apply different rationales to the same practice. In a bid to portray some of the complexity and tensions, and how the three teachers constructed their use of the term AfL, we experimented with different ways of conveying the findings. In part inspired by the approaches in Dowling, Fitzgerald, and Flintoff (2012), we agreed to retell an assessment story for each teacher based on their interactions throughout the focus groups. The stories are specifically structured to reflect AfL strategies shared in Table 1, and we recognize the constructed assessment stories as part of the analysis. They could be 'true' in the sense that they are generated from real events and teachers' conversations, and we quote passages from the respective focus groups and try to keep the vocabulary and syntax close to that of the physical education teacher. The intention is to move the focus from information collection and reporting of what the teachers share to encouraging the readers to revisit and rethink their understanding and beliefs about learning and assessment practices. It is anticipated that the constructed assessment stories will allow us to consider the extent to which teachers' references to AfL align with AfL theory and discuss how the stories relate to the wider contexts of physical education and movement culture.

### **Leif's story – curious and prepared to experiment with AfL strategies**

Leif is in his late thirties and has been teaching physical education for eight years. Leif works with Ronald at a relatively new upper secondary school in Oslo and they have been colleagues for two years. With approximately 900 students at the school (boys and girls aged 15–19 years), there are four other physical education teachers at the school.

Leif reports the level of support from the school leadership of the new school as disappointing. Without a clear school policy for assessment, the physical education teachers have attempted to agree a common structure (four main teaching units of discrete content over the year) for physical education. Each unit ends with a 'test', a planned assessment situation where the students have to demonstrate knowledge and practical skills. In attempting to explain what this new and different discourse on assessment in physical education entailed, Leif commented:

One important aspect was that they [students] would meet these questions again at the end of the year, exactly the same. This way we tried to make it assessment for learning. They knew (...) at some point they would get the same questions again, so if I want to do better, I can check my old test and assure that I am capable of answering what I was supposed to back then. This way it was supposed to be assessment for learning, that was the thought behind it all.

Leif's explanation is an example of the formative use of a summative test, implying that the connection he makes to AfL is meaningful. Throughout the focus group, Leif communicated commitment to accommodating and arguing for what he referred to as 'skill-tests' (i.e. assessments that measure students' ability to perform a skill within a discrete drill). Due to the expectation of the skill tests informing the students where they are and where they are going next, Leif first not only presents them as an attempt to implement formative assessment practice but also states that 'the skill tests were carried out mainly because we were required to'. In this instance, the engagement in developing assessment becomes primarily a task of accountability and the requirements of documentation, rather than focusing towards how to use assessment to improve students' learning.

As the focus group develops, Leif extends the argument that there are more important aspects to assessment than solely assessing a student's ability at performing a particular skill. He explains how the physical education teachers try to include self-assessment, one of the main AfL strategies, as part of the skill tests and states that 'the fun part is that the students almost always have put down the same grade as me'. Leif admits that he struggles to be sufficiently precise to allow him and the students to share 'the same perception of what the grade 4 look like and when the student has arrived at 4'. The examples of self-assessment provided by Leif are all concerned with grading, resulting in him describing summative assessment practice.

Leif is also a qualified mathematics teacher, and makes a comparison of assessment in the two subjects he teaches:

I teach mathematics too, and I think the assessment contexts are two opposites. In mathematics I can sit down at the table and a cup of coffee and a sheet, so I can sit and consider; 'that's correct', 'that's wrong', and then I count up and get to 97% correct answers which give grade 6. (...) Physical education is the subject of assessing moments, or 'short-cuts', right? Twenty seconds, I observed Lisa who did some skilled and entertaining moves. Then something happens in the changing rooms which needed my attention. Later, I ask; 'hell, what just happened there?' What was it that Lisa did, was the short moment showing an important side of her movement competence?' Finding a system then, which systematises and records such short moments which in the end also is supposed to constitute the final grade.

Comparing assessment across the subjects, Leif highlights the complexity of the teaching and learning contexts and assessment challenges in physical education, noting that what is at the core of assessment in each school subject is presented differently. While ongoing assessment has a prominent role in physical education, the focus in mathematics appears to be concerned with arriving at a final grade. While Lisa 'doing some skilled and entertaining moves' during one lesson can influence assessment of Lisa in physical education, opportunities to convey learning in less-structured learning contexts are not accommodated in mathematics.

Leif is aware of the concerns related to skill tests as planned assessment situations and highlights the importance of understanding how continual assessment can be used to inform students' learning:

I just want to make a little comment to the on-going assessment, as it was also something I did make clear, that if you do not have a good day at the skill-tests, (...) and make a much worse impression than you are able to, then I will know. I will know from how you participate and perform on activities during the weekly lessons, and that overrides what you do on a test. I tried to be very clear on this issue, although students may not have seen it that way.

In concluding, while Leif expresses a continuous curiosity and will to continue experimenting with assessment, his practice also conveys a lack of systematic reflection, and perhaps a vocabulary, to allow him to interrogate effective assessment further.

### **Ronald's story – tensions between enacting different assessment cultures**

Ronald teaches in the same school as Leif, is in his late fifties and has been a full-time physical education teacher in upper secondary school for the last 20 years. Ronald is informed about AfL and positions himself somewhat differently from Leif in the ways in which he has attempted to implement AfL into practice. From the outset, Ronald appears to be an experienced physical education teacher who has recently found inspiration to explore new approaches to assessment. He admits that the previous year was the first time he did not complete a pen and paper test with the physical education students and, 'in relation to assessment which enhance learning, or assessment for learning, (...) I tried out some things I have not done before, for

example, that students are being engaged in self-assessment'. Ronald continues by providing an example of what this can look like:

After finishing a task, for example, they [students] as a group sit down and put down some points. What was good? What could have been better? Then I can do the same, and then we can talk about it. And so, not to agree on a grade, but we both put down a suggestion for a grade. So we'll see if it matches, (...) if I – who, of course always assesses correctly (laughter) – have graded a four and the student have six, so there we at least have something to talk about, and if we both have, for example five, then yes, we might suggest that it must be quite correct.

When it comes to clarifying criteria and providing feedback in accordance with the competence aims in the physical education curriculum, Ronald is supportive of the structure commenting that learning goals related to the competency aims heighten student learning. Ronald argues that it is easier to provide feedback and tutor students towards achieving competence aims when learning goals are more manageable and concrete. He also makes reference to the practical nature of physical education and how, rather than rely on providing oral feedback to the students, teachers will 'prepare a practical learning experience that makes them progress'. Ronald does admit to the inherent challenge in verbally communicating to students the connection between competence aims and enacted learning experiences.

Ronald is also conscious about assessment and grading in physical education being a way to communicate what is to be regarded as important in the subject, particularly when the 2006 school reform explicitly articulated that the level of student effort should not influence grading in the subject. Ronald does not agree with this and admits to accommodating level of effort in his assessments:

Halfway through the semester I noted down an assessment, which I called participation and activity level, on each student, and I did the same again at the end of the semester. This is to ensure that this aspect also would be considered, as well as pen and paper tests and workbook, how fast they were able to run and how they performed in basketball, floorball and stuff like that.

Reflecting the criterion-referenced and decentralized assessment system in Norway, Ronald explained how he also has rethought the concept of fairness:

Ronald: In terms of what can be difficult by student assessment (laughter) but at least, according to how my understanding gradually has developed, fairness is a challenge. Some might have the view that fair would be if everything was equal, a benchmark to assess for, some think that's fair, right?

Researcher: Yes?

Ronald: But the way I have come to understand it, eventually, is well that fairness could also be that students would get a clue of the overall aims of our practice and the assessment procedures and that this is prepared in advance so that they have something to relate to. Whether it is one or the other, it will vary between schools and classes, and it will, because that's the system now, it varies yes, because we here at our school determine what is fair and a teacher and student at another school

determine what they consider to be fair. In this way the crucial point is that they are aware of what is going to happen and what the criteria for success are.

Ronald has made space to engage students in assessment and explore, through his own practice, how to integrate AfL strategies into teaching and learning. He is an experienced teacher motivated to change assessment practice, believing that assessment can become more effective and make an important contribution to student learning in physical education. He does admit that he, students and some colleagues are still drawn at times to the traditional assessment culture where the allocation of a final grade has dominated practice. Admitting that he had not encouraged or facilitated peer-assessment in his classes, Ronald expresses this tension between the different assessment cultures or traditions:

I have not used peer-assessment, but I remember I asked [the students] about it, and the reaction to it was relatively negative. However, the frame was that they should put a grade on each other's performance or competence. It's very easy when we talk about assessment, to think grading, and start discussing grading. But if we really mean an assessment, more like guidance, then it, of course, becomes another matter. Then I think peer-assessment will be ok, in for example basketball and whatever else we do. It will be very good for students to tutor each other. But, you know, we do not often think about it that way, because when we talk about assessment we talk about numbers.

Laughing, he adds that 'it's really misunderstood, because it's the other assessment that is the key, really, to promote learning, right?'

### **Geir's story – an evolving disposition towards formative assessment**

Geir is in his late thirties and a full-time physical education teacher of 14 years who works in an upper secondary school in Oslo. The school has a relatively small student intake of approximately 300 boys and girls aged between 15 and 19 years. There are two physical education teachers in the school and consequently the focus group that Geir participated in included only his female colleague and the researcher. The physical education programme at the school entails a broad range of activities, mostly sport orientated. Geir and his colleague express that the level of support from the school leadership is good, and the school has participated in a national project for developing assessment practice (Utdanningsdirektoratet, 2011). They also note that they have good informal opportunities to discuss and reflect with each other on assessment practices.

Geir is quick to allude to the tension with the sole use of a 'skill track' (similar to what Leif had termed a 'skill-test') to encompass and demonstrate learning and the concept of ongoing assessment promoted by AfL:

We practice (...) the basic skills and then the student can test their skills in a skill track at the end (...) this is the formal assessment situation. (...) But I try to tell students that I also do a more comprehensive assessment [than that of a skill track/



skill test]. So that you may well have a bad day on test, if one has shown earlier in rehearsal lessons that he or she possesses the skills, I take that into account. It's a bit like that in relation to assessment for learning, that we are concerned with practice and rehearsal prior to testing.

While Geir appears to favour a skill track as a feasible assessment for physical education, he does provide instances where he engages with a more eloquent understanding of AfL when prompted on considering the relationship between the awarding of a grade and feedback:

What's important, also relating to feedback, is that the students know what is expected of them, that they know what they are going to be measured in relation to. So in advance, at the beginning of a period, we present the competence aim, and so, at least I try to break down the competence aim into learning objectives, then, it all becomes a little more concrete for students. (...) And so we present assessment criteria, what they must convey to get the different grades. Sometimes this is presented orally. Sometimes they get it on paper. (...) And then, during practice it is all along about giving them feedback on what they do well and things they can do better.

Geir explains how his teaching is somewhat guided by the 'two stars and a wish' concept. A literal translation of the concept is to acknowledge at least two things that the student is doing well before introducing a focus on something they can work on to improve their performance. Geir shares an evolving disposition towards formative assessment when he alludes to introducing self-assessment, admitting that he is still exploring the concept. He has used a 'thumbs up, thumbs down' technique to garner from students their assessment of their own experience in a task (i.e. thumbs up meaning they are pleased with their performance of the task and thumbs down that they are not pleased). Geir is conscious that student self-assessment is not solely for the benefit of the teacher to assess how individual students are reacting to a task but rather that it works to the benefit of the student, 'There is more about awareness for their own sake, and I think, that's a little of what self-assessment is all about'. Geir also admits to only recently introducing peer-assessment and notes the extent to which he is happy with how the students have reacted to such involvement.

The importance of sharing criteria for success is inadvertently mentioned by Geir when he comments that he believes that those students who are dissatisfied with their skill track grade are those students who did not appreciate what the assessment criteria was, believing they have invested sufficient effort to warrant a better grade. Geir goes on to consider ways in which physical education should assess effort more comprehensively, believing that 'effort, and being active and participating is central to promote more physical activity, and health and lifestyle'. He considers how assessment in physical education can best accommodate what the subject area is/should be concerned with promoting:

If we were to talk about what one would wish that physical education should be then. What role it could play in the school? So, communicating the importance of

being physically active is part of it. To understand how you can benefit from it, and also to motivate the interest of sports and activities. One idea is that the students can be exposed to some activities which they would like to participate in or do on their own. But certainly, there must be knowledge and skills as well.

Geir alludes to the tensions that arise when he explains the value of skill tracks to students who question the appropriateness of such assessments when, dependent on the skill, students can score highly in the assessment due to already being proficient performers in specific skills. Interesting, it is the students who appear to be requesting a reconsideration of the appropriateness of skill tracks in this instance and not the teacher. Somewhat contradictory to what Geir alluded to previously, with respect to what physical education should be acknowledged as promoting, Geir advises students that it is not enough for them to show up in class and participate but rather that they need to be able to perform a skill at a certain level to achieve a respectable grade, advising students:

So between lessons, even if you do not get it specified as homework you need to go home and [practice]. They must understand that according to skills, well, it's not enough to be present and to show much good effort.

Geir's focus on sharing the criteria for success, and few other components of AfL, appears to be somewhat responsible for his lack of engagement in considering assessments that encourage and acknowledge student investment and effort in a task. Geir believes that if students convey a level of effort it is likely that their skill set will improve.

Geir believes that the final assessment point in physical education in his school has improved due to a more systematic approach to formulating assessment criteria and clarifying how learning is to be assessed, both of which are identified as AfL strategies. However, he appears unable to believe that 'students are ready to deal with the abstract level of the competence aims' and subsequently continues to rely on the teaching of intricate and discrete skills.

## **Discussion**

Due to the selection criteria for the three teachers involved in this study (i.e. they had, unprompted, referred to AfL in focus groups concerned with assessment), it was not surprising that all three teachers revealed a high level of engagement and reflection with assessment, and the associated tensions in changing their own assessment preferences and practices. Throughout the focus groups, they provided descriptions of engagement with all five key strategies of AfL (William & Thompson, 2008). It is not clear whether they recognized a more general need for change in physical education (Kirk, 2010; Säfvenbom et al., 2014), or if their engagement with AfL stems mostly from what they are expected to do according to official requirements and new assessment regulations as set out in the previously mentioned Norwegian school reforms.

Recalling the double-edged sword metaphor about the mandatory aspects of physical education (Säfvenbom et al., 2014), it is somewhat surprising that analysis of the focus group interactions did not reveal critical reflection about how assessment practice can, to follow the metaphor, end up on the 'wrong' side of the sword. In other words, the teachers did not reflect upon how assessment can have unintended, and for some students harmful, consequences that may contribute to negative attitudes towards movement contexts in general. This indicates that the physical education teachers need to develop awareness of assessment as a process in which distribution of power is inherently disproportional, and challenge the 'naturalness' of assessment practices, performances and outcomes (Hay & Penney, 2013).

AfL strategies and the assessment stories of Leif, Ronald and Geir lead us to focus on engaging the students in a more active role as learners in physical education. This is one way of empowering students in regard to assessment and is potentially an effective way of linking movement and learning in physical education to their life as 'movement consumers' outside school. For example, teachers can facilitate learning experiences where students are invited to present and discuss different movement contexts that they already are part of. Geir, Ronald and Leif all alluded to the value of encouraging students to engage with where they currently reside in their learning and to pursue the goals that they are striving to achieve. However, conscious of the limited amount of self- and peer-assessment, it becomes clear that even Geir, who appeared to have engaged more fully with such strategies, admitted that he was still in an exploratory phase when it comes to promoting such forms of assessment. Leif and Ronald seemed to engage with self-assessment only at the end of the unit when they asked students how they would grade themselves. This could be construed with the popular notion that teachers consider assessment synonymous with providing a final grade.

Understanding the distinction between the processes of assessment and grading is an important part of teachers' assessment literacy (Hay & Penney, 2013), but it also resonates an unsolved tension between the summative and formative purposes of assessment at the national policy level. All three teachers conveyed an appreciation of formative assessment's contributions to summative assessment, acknowledging that it was not necessary for the final 'skill-tests' performance to be solely dependent on the enactment of a summative assessment. There were a number of instances where AfL strategies were compromised due to the external requirements of assessment. Leif explained that he sometimes completed the feedback component of AfL more from his obligation to address the national curriculum requirements than as a mechanism through which to encourage students' own learning. The broader structures surrounding assessment become important if we are to understand such tensions and ambiguous descriptions regarding AfL in the assessment stories, appreciating that there was little support at the teacher level following the 2006 reform (Møller & Skedsmo, 2013). In addition, physical education teachers motivated to develop their formative assessment practices faced an almost complete lack of examples or descriptions of how to do AfL in movement contexts (Georgakis & Wilson, 2012). Teachers were, as conveyed in our assessment stories, predetermined

to explore, rather than implement, AfL in physical education. In this regard Geir, Ronald and Leif can be interpreted as reflective professionals trying to move away from physical education as 'a field where other abilities are valued than the ones to be found in the syllabus' (Redelius et al., 2009, p. 259). From an educational research perspective, we appreciate that they place learning at the foreground and will use assessment to clarify what is to be valued in the subject. At the same time, we are concerned to observe how certain practices developed in the name of AfL actually narrow the concept of learning, mostly under the influence of a still remaining 'meritocratic mode' (Crum, 1993), or in Kirk's (2010) terms 'physical education-as-sport-techniques' (p. 4).

The teachers' attempt to enact and interpret AfL appears to be constricted to the traditional emphasis in Norwegian physical education on linking assessment to classifying and grading students, as well as feeling pressurized to adjust practices to official requirements concerning documentation and accountability (Arnesen et al., 2013). In some cases, the teachers' commitment to clarifying and sharing criteria on behalf of the national curriculum seemed to be more important than planning and teaching worthwhile, meaningful and relevant student learning experiences. To a degree, it appears that the notion of 'learning' in physical education is confused with 'criteria compliance'. If so, the potential of AfL as being 'oriented towards greater equity within and beyond physical education' (Hay & Penney, 2013, p. 110) vanishes, replaced by what looks like assessment for assessment's sake.

### **Concluding remarks**

This paper has shared the stories of three Norwegian physical education teachers exploring 'alternative' assessment practices which they identify as AfL. All three teachers appeared to be well informed of the national curriculum, assessment regulations and the purposes of implementing AfL. However, it was evident that their understanding and enactment of AfL key strategies was somewhat constricted. If the physical education profession is unable to find ways in which to address this, we risk not supporting assessment reform with tragic consequences for student learning. It is therefore crucial that physical education teachers focus on the individual learner and appropriate learning experiences. There is a need of thorough teacher planning, preferable with systemic moderation and support for students to take on new roles and responsibilities in their own learning. Recalling Bennett (2011), we reiterate that intellectual and instructional tools 'ought to be specifically tuned for the domain in question' (p. 16), and hope more stories regarding physical education practices can help us figure out what this will look like.

We are concerned to observe in this instance that no real effort appears to have been made to create the supportive environment in physical education for successful implementation of the AfL strategies. Our concern is underpinned by research pointing to teacher professional development as a prerequisite for successful implementation of AfL (Smith, 2011), and the need of embedding AfL in learning theory (Gardner, 2012). This may well be a challenge, acknowledging

that not only are most existing theories of learning defined cognitively, but also that learning connected to physical education and activity is, to a large extent, practical and embodied, and also linked to the powerful discourses of sport and related areas such as health (Quennerstedt et al., 2014). Similar to Hay (2006) and López-Pastor et al. (2013), and acknowledging the limitation of a study on three physical education teachers, we advocate the need for more research on AfL related to physical education. It is only through further exploration that we will begin to track the extent to which AfL, as an integrated part of pedagogical practice, can contribute to an idea of physical education as a 'planned introduction to movement culture' (Crum, 1993, p. 352).

### Acknowledgements

We are very grateful to the two anonymous reviewers for focused and constructive comments on the original version of the paper.

### References

- Alvesson, M., & Skoldberg, K. (2009). *Reflexive methodology: New vistas for qualitative research*. London: SAGE.
- Annerstedt, C. (2008). Physical education in Scandinavia with a focus on Sweden: A comparative perspective. *Physical Education and Sport Pedagogy*, 13, 303–318. doi:10.1080/17408980802353347
- Annerstedt, C., & Larsson, S. (2010). 'I have my own picture of what the demands are ...': Grading in Swedish PEH – problems of validity, comparability and fairness. *European Physical Education Review*, 16(2), 97–115. doi:10.1177/1356336X10381299
- Arnesen, T. E., Nilsen, A. K., & Leirhaug, P. E. (2013). 'Den læreplanen som ikkje kan tilpassast mi undervisning, finst ikkje.' Vurdering og undervisning i kroppøving etter Kunnskapsløftet [Assessment and teaching in physical education after the reform the Knowledge Promotion]. *Tidsskriftet FOU i praksis*, 7(3), 9–32.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved from [http://assessmentreformgroup.files.wordpress.com/2012/01/10principles\\_english.pdf](http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf)
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5–25.
- Bernstein, B. (1975). *Class, codes and control, volume 3: Towards a theory of educational Transmission*. London: Routledge & Kegan Paul.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–73.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Education, Assessment, Evaluation and Accountability*, 21, 5–31.
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse of assessment in physical education. *Asia-Pacific Journal of Health, Sport and Physical Education*, 2(1), 3–19.
- Crum, B. J. (1993). Conventional thought and practice in physical education: Problems of teaching and implications for change. *Quest*, 45, 339–356. doi:10.1080/00336297.1993.10484092
- Dowling, F., Fitzgerald, H., & Flintoff, A. (2012). *Equity and difference in physical education, youth sport and health: A narrative approach*. London: Routledge.
- Gardner, J. (Ed.). (2012). *Assessment and learning*. London: SAGE.
- Georgakis, S., & Wilson, R. (2012). Australian physical education and school sport: An exploration into contemporary assessment. *Asian Journal of Exercise & Sport Science*, 9(1), 37–52.

- Green, K. (2014). Mission impossible? Reflecting upon the relationship between physical education, youth sport and lifelong participation. *Sport, Education and Society*, 19, 357–375.
- Hay, P. (2006). Assessment for learning in physical education. In D. Kirk, D. Macdonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 312–325). London: SAGE.
- Hay, P., & Macdonald, D. (2008). (Mis)appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education*, 15, 153–168. doi:10.1080/09695940802164184
- Hay, P., & Penney, D. (2013). *Assessment in physical education. A sociocultural perspective*. New York: Routledge.
- James, A. R., Griffin, L., & Dodd, P. (2009). Perceptions of middle school assessment: An ecological view. *Physical Education and Sport Pedagogy*, 14, 323–334. doi:10.1080/17408980802225792
- Kirk, D. (2010). *Physical education futures*. London: Routledge.
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18, 57–76.
- MacPhail, A., & Halbert, J. (2010). 'We had to do intelligent thinking during recent PE': Students' and teachers' experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy and Practice*, 17(1), 23–39.
- Møller, J., & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Education Administration and History*, 45, 336–352.
- Ni Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18, 219–233.
- Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research and Method in Education*, 29(1), 23–37. doi:10.1080/01406720500537304
- Quennerstedt, M. (2013). Practical epistemologies in physical education practice. *Sport, Education and Society*, 18, 311–333.
- Quennerstedt, M., Öhman, M., & Armour, K. (2014). Sport and exercise pedagogy and questions about learning. *Sport, Education and Society*, 19, 885–898. doi:10.1080/13573322.2013.847824
- Redelius, K., Fagrell, B., & Larsson, H. (2009). Symbolic capital in physical education: To be, to do or to know? That is the gendered question. *Sport, Education and Society*, 14, 245–260.
- Redelius, K., & Hay, P. (2009). Defining, acquiring and transacting cultural capital through assessment in physical education. *European Physical Education Review*, 15, 275–294. doi:10.1177/1356336X09364719
- Säfvenbom, R., Haugen, T., & Bulie, M. (2014). Attitudes toward and motivation for PE. Who collects the benefits of the subject. *Physical Education and Sport Pedagogy*. Advance online publication. doi:10.1080/17408989.2014.892063
- Scriven, M. (1967). The methodology of evaluation. In P. A. Taylor & D. M. Cowley (Eds.), *Reading in curriculum evaluation* (pp. 39–83). Dubuque, IA: Wm. C. Brown.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. doi:10.3102/0013189X029007004
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. doi:10.3102/0013189X015002004
- Smith, K. (2011). Professional development of teachers – A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation*, 37, 55–61. doi:10.1016/j.stueduc.2011.03.005

- Tannehill, D., van der Mars, H., & MacPhail, A. (2013). *Building effective physical education programs*. Sudbury, MA: Jones & Bartlett.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Philadelphia, PA: Open University Press.
- Utdanningsdirektoratet. (2011). *Grunnlagsdokument: Satsingen Vurdering for læring 2010–2014* [Background document: Assessment for learning 2010–2014]. Oslo: Author. Retrieved from <http://www.udir.no/PageFiles/Vurdering%20for%20laring/Dokumenter/Nasjonal%20satsing/2/Grunnlagsdokument%20for%20satsingen%20Vurdering%20for%201%c3%a6ring%20okt%20%2020111.pdf>
- Ward, G. (2014). Learning movement culture: Mapping the landscape between physical education and school sport. *Sport, Education and Society*, 19, 569–604. doi:10.1080/13573322.2012.690342
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14. doi:10.1016/j.stueduc.2011.03.001
- Wiliam, D., & Thompson, A. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum.





Vedlegg 1:

Informasjonsskriv med forespørrelse til elever om deltaking i spørreundersøkelsen



Invitasjon til spørreundersøkelse for elever i prosjektet "vurdering for læring i kroppsøving?"

Les [hele teksten på dette arket](#) for elevene før de logger inn på nettstedet <https://resp.nsd.no> ved hjelp av tildelt innloggingsID og PINkode.



## **Forespørsel om deltaking i spørreundersøkelse i forbindelse med forskningsprosjektet "Vurdering for læring i kroppsøving?"**

I arbeidet med min doktorgrad studerer jeg læring og vurdering i kroppsøving ved flere videregående skoler. Skolen du er elev ved er en av skolene som har sagt seg villig til å være med i prosjektet. Prosjektet består blant annet av gruppeintervju med lærere som underviser i kroppsøving og en nettbasert spørreundersøkelse blant alle elever ved skolene.

Hensikten med spørreundersøkelsen er å få fram ulike oppfatninger og erfaringer med innhold, vurdering og læring i kroppsøving. Spørsmålene handler om hva dere gjør og opplever i kroppsøving, hva dere forventes å lære i faget og hvordan vurderingen foregår. Du skal svare ut fra dine opplevelser med faget på din skole. I tillegg vil du få noen spørsmål om din bakgrunn og aktivitet på fritida.

Alle opplysninger som samles inn vil bli behandlet konfidensielt. Både skole og enkeltpersoner anonymiseres ved formidling/rapportering av resultater. Spørsmålskategorier med svar fra mindre enn fem elever vil grupperes slik at de ikke kan knyttes til den enkelte skole. Fullstendig anonymisering vil gjøres etter eventuelle oppfølgingsstudier, senest i 2020. Prosjektet er godkjent av Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste A/S.

Det er frivillig å være med og du har mulighet til å trekke deg når som helst underveis, uten å måtte begrunne dette nærmere. Dersom du trekker deg vil alle innsamlede data om deg slettes.

Ved å logge inn og svare på undersøkelsen bekrefter du å ha mottatt denne informasjonen om forskningsprosjektet *Vurdering for læring i kroppsøving?* og at du er villig til å delta i spørreundersøkelsen.

Med vennlig hilsen  
Petter Erik Leirhaug  
Stipendiat ved Norges idrettshøgskole



Vedlegg 2:

Papirversjon av studiens internettbaserte spørreskjema



**Les dette før du starter:**

*Du skal nå ha fått opplest en forespørsel om å delta i denne spørreundersøkelsen om kroppsøving, der hensikten er å undersøke ulike oppfatninger og erfaringer med innhold, vurdering og læring i kroppsøving. Ved å klikke deg videre bekrefter du å ha mottatt denne informasjonen og at du er villig til å delta.*

*Hele undersøkelsen består av 32 hovedspørsmål. Det tar omtrent 20 minutter å gjennomføre undersøkelsen.*

*Det er viktig for resultatet at du svarer så ærlig og godt som mulig. Du skal svare ut fra dine opplevelser med kroppsøving på din skole. Husk at ingen svar er mer riktige enn andre og ingen får vite hva du har svart.*

*Lykke til!*

---

### Kjønn

- Jente  
 Gutt
- 

### Hvilket trinn går du på?

- Vg1  
 Vg2  
 Vg3
- 

### Hvordan er din innstilling til å ha følgende i kroppsøving?

	1 Svært negativ	2	3	4	5 Svært positiv
Konkurranser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Egentrening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hjemmearbeid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teoritimer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Karakterer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skriftlige prøver	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Praktiske prøver	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å selv få lede/instruere	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Styrketester	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Utholdenhetstester	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aktiviteter der jenter og gutter er hver for seg	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

### Hvor mye beveger du deg i kroppsøvingstimene?

- Jeg er aldri eller nesten aldri med i kroppsøvingstimene  
 Jeg deltar bare noen ganger i kroppsøvingstimene  
 Jeg er nesten alltid med, men jeg beveger meg ikke spesielt mye  
 Jeg er nesten alltid med, og jeg beveger meg mye
-



### Hvor ofte har du i løpet av det siste året i kroppsøving eller på aktivitetsdager hatt disse aktivitetene?

	Aldri hatt	1-2 ganger	3-4 ganger	5-10 ganger	Mer enn 10 ganger
Dans du selv er med på å lage	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Egentrening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rugby	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Basketball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Volleyball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rekkertspill (f. eks. badminton, tennis)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fotball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Handball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Innebandy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tester (av ulike slag)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Friluftsliv (f.eks. tur i skogen, matlaging ute)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Styrketrening med vekter/treningsapparater	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Friidrett (f. eks. lenghe hopp, høyde, baneløp, kulestøt)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Orientering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Og noen aktiviteter til: Hvor ofte har du i løpet av det siste året hatt disse?

	Aldri hatt	1-2 ganger	3-4 ganger	5-10 ganger	Mer enn 10 ganger
Leik (og annen bevegelsesmoro)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Svømming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Klatring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Padling i kano/kajakk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Turn/akrobatikk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trampoline eller trampett	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aerobic og andre former for musikkmosjon	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kampsport (f. eks. judo, karate, capoeira)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Norsk folkedans (f. eks. reinlender, polka, halling)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dans fra andre kulturer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skigåing – langrenn, tur eller skileik	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ski, snøbrett, eller liknende i alpinbakke	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skøyter/isspill	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intervalltrening for utholdenhet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Yoga, Pilates, massasje eller liknende avspenningstrening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Har dere hatt undervisning der kroppsøving har vært slått sammen med andre fag (tverrfaglig undervisning)?

- Ja
- Nei
- Vet ikke

**Kjenner du kompetansemålene i kroppsøving? (sett et kryss der det passer best)**

- Nei, de er jeg ukjent med
- Ja, jeg har lest dem på egen hånd
- Ja, læreren har gått gjennom dem
- Ja, jeg har lest dem og læreren har gått gjennom dem
- Ja, vi arbeider stadig med kompetansemålene

**Hvor viktig mener du kroppsøving er i forhold til de andre fagene på skolen?**

- Mindre viktig enn andre fag
- Like viktig som andre fag
- Viktigere enn andre fag

**Hvor mye betyr følgende for vurderingen i kroppsøving ved din skole?**

	1 Betyr ikke noe	2	3	4	5 Betyr svært mye
God utholdenhet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å være fysisk sterk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teoriprøver	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Innsats i timene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lite fravær	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kompetanse i friluftsliv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gode ferdigheter i ballspill	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gode ferdigheter i dans	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
At du er god i en enkelt idrett	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å dusje	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å vise framgang (f. eks. i styrke, utholdenhet og ferdighet)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å trene på fritida	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Plan for egentrening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gjennomføring av egentrening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Måloppnåelse (kompetansemål)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
At du er bedre enn andre elever i idrettene/aktivitetene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gode prestasjoner i konkurranseidrett (utenfor skolen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Hvilken karakter fikk du i kroppsøving ved siste halvårsvurdering?**

- 1
- 2
- 3
- 4
- 5
- 6
- Fikk ikke karakter

**Hvor godt passer disse påstandene med din oppfatning av karakterer og vurdering i kroppsøving?**

	1 Helt uenig	2	3	4	5 Helt enig
Kroppsøvlingslærer samtaler med meg om karakteren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg synes jeg får den karakter jeg fortjener	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
For å få god karakter må jeg gjøre det bra på tester	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Elevene er med og vurderer hverandre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er fornøgd med kroppsøvlingskarakteren min	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kroppsøvlingslæreren gir rettferdige karakterer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg deltar i kroppsøving bare for å få karakter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vurdering handler mer om hva jeg skal lære enn om karakter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jenter har lettere for å få god karakter i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kroppsøvlingslærerne på skolen bruker karakterene likt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg vet hva som kreves for å få de ulike karakterene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kroppsøvlingslæreren har forklart om vurdering for læring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg får være med å vurdere meg selv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg vet hva jeg må gjøre for å få bedre karakter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg kunne fått bedre karakter i en annen klassegruppe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ved vår skole er kompetansemålene grunnlaget for vurdering i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Har du i enkelttimer eller deler av kroppsøvlingsfaget hatt undervisning eller eget opplegg som har vært spesielt tilrettelagt for deg?**

- Ja  
 Nei

**Har du noen funksjonshemming eller kronisk sykdom som gjør det vanskelig å delta i (deler av) kroppsøvlingsfaget slik det blir drevet ved din skole?**

- Ja  
 Nei

## Hvordan passer disse påstandene med din selvoppfatning?

	1 Helt uenig	2	3	4	5 Helt enig
Jeg trives godt på skolen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg opplever ofte mestring i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg har gode venner i klassen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er fornøyd med kroppen min	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er fornøyd med mine prestasjoner i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Det er greit å skifte og dusje på skolen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg føler meg ofte utrygg i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg prøver å vise god orden og oppførsel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er i dårlig fysisk form og blir fort sliten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er generelt skolelei	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er motivert for å trene og lære nye ferdigheter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kroppsøving er det dårligste faget mitt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg trives best i kroppsøving når vi arbeider/trener individuelt eller i par	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg liker å vise meg fram i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg liker å trene/stå på så jeg blir skikkelig fysisk sliten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg ønsker å ha en mer aktiv livsstil enn jeg har	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

### Hvor godt passer disse påstandene med din oppfatning av kroppsøving?

	1 Helt uenig	2	3	4	5 Helt enig
Jeg lærer meg hva som skjer i kroppen ved oppvarming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg forbedrer mine ferdigheter i ulike idretter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer meg å samarbeide med andre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer ikke noe nytt i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg får avkopling fra det vanlige skolearbeidet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg har det gøy i nesten alle timene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Noen elever i klassen deltar nesten aldri i timene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg skulle helst sluppet å være med	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg synes det er kjekt når vi deler inn lag	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer hvordan fysisk aktivitet kan påvirke helsa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg blir aldri sliten i timene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer å planlegge trening og trene på egen hånd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teori i kroppsøving er unødvendig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg gjør så godt jeg kan i kroppsøvingstimene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg synes vi skulle hatt flere timer kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer hvordan lage bål og mat ute	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer om kosthold og helse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer om idrettsskader og førstehjelp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer ulike danser	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg lærer fair play	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Omtrent hvor mange ganger i dette skoleåret har dere i kroppsøving snakket om kompetansemålene?

- Aldri
- 1 gang
- Ca. 2-4 ganger
- Ca. 5-10 ganger
- Over 10 ganger

### Omtrent hvor mange ganger i dette skoleåret har du vært med å vurdert ditt eget arbeid i kroppsøving?

- Aldri
- 1 gang
- Ca. 2-4 ganger
- Ca. 5-10 ganger
- Over 10 ganger

**Omtrent hvor mange ganger så langt i dette skoleåret har du fått tilbakemelding fra lærer om hva du må gjøre for å forbedre kompetansen din i faget?**

- Aldri
- 1 gang
- Ca. 2-4 ganger
- Ca. 5-10 ganger
- Over 10 ganger

**På hvilke måter får du tilbakemelding fra kroppsøvlingslærer?**

*(Her kan du sette flere kryss).*

- Tilbakemelding underveis i timene, ved at lærer hjelper til fysisk/viser meg
- Tilbakemelding underveis i timene, ved at lærer gir muntlige råd og tips til hva jeg bør prøve/gjøre/trene vider på
- Tilbakemelding i på forhånd avtalt samtale
- Tilbakemelding i samtale der lærer tar deg til side i en time
- Tilbakemelding på eget skjema for tilbakemelding
- Tilbakemelding på Fronter / itslearning
- Tilbakemelding på andre måter

**Hvor godt passer disse påstandene med din oppfatning av undervisningen i kroppsøving?**

	1 Helt uenig	2	3	4	5 Helt enig
Jeg får hjelp når jeg trenger det	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vi har skriftlig prøve en gang i halvåret	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vi har ballspill nesten hver time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vi følger utdelte periodeplaner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Det er mye bråk og uorden i kroppsøvingstimene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg får tilbakemeldinger om hvor jeg er i læreprosessen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg blir ofte nødt til å gjøre ting jeg ikke liker	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Læreren har god oversikt over hva jeg kan og ikke kan	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vi har for mye dans	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vi har for mye friluftsliv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg får ofte støtte og hjelp av mine medelever	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg blir motivert til å trene og være aktiv i hverdagen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Innholdet i timene er lite variert	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Det er for lite hard trening i kroppsøvingstimene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lærer forskjellsbehandler elever	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lærer deltar selv ofte i idrettene/aktivitetene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lærer er flink til å motivere og engasjere elevene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vi har aldri hjemmearbeid (lekser) i kroppsøving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Egenvurdering brukes systematisk (trenger ikke være ofte)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Alle elevene er i aktivitet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Hvilket kjønn er kroppsøvingslæreren din?

- Kvinne
  - Mann
  - Har flere lærere, begge kjønn
- 

### Hvilken oppvekstbakgrunn har du?

- Jeg er født og oppvokst i Norge
  - Jeg er adoptert, oppvokst i Norge
  - Jeg har innvandringsbakgrunn (født i annet land)
  - Jeg har innvandringsbakgrunn (født i Norge, mine foreldre/foresatte innvandret)
- 

### Hva er din mors høyeste utdanning?

- Høyskole/universitet
  - Videregående
  - Grunnskole
  - Vet ikke
- 

### Hva er din fars høyeste utdanning?

- Høyskole/universitet
  - Videregående
  - Grunnskole
  - Vet ikke
- 

### Hvor ofte har du i snitt per uke (i fritida) drevet med følgende treningsaktivitet i løpet av de 12 siste måneder?

	Aldri	Under 1 gang p/uke	1 gang p/uke	Flere ganger p/uke
Utholdenhetsidrett (feks løp, sykling, langrenn, svømming)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lag-/ballidretter (feks squash, håndball, fotball, ishockey)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Styrkeidrett (feks bryting, vekttrening)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kampsport (feks judo, karate, taekwondo)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tekniske idretter (feks ridning, alpint, telemark, friidrett, snowboard, golf, rullebrett/skøyter)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Risikoidrett (feks elvepadling, fjellklatring, paragliding)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Treningssenter (feks spinning, saltrening, apparatsal)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dans (feks hip-hop, ballett, leikarring, salsa)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

### Hvor ofte, i en normal uke, er du sammenhengende fysisk aktiv med trening, rask gange, dans, idrett eller annen aktivitet i mer enn en 1/2-time?

- Hver dag
  - 5-6 dager per uke
  - 3-4 dager per uke
  - 1-2 dager per uke
  - Veldig sjelden
-

**Konkurrerer du i idrett, dans eller lignende? Både individuelle konkurranser og der du deltar på et lag teller med.**

- Ja  
 Nei

**Hvor mange ganger totalt har du i løpet av de siste 12 måneder gjort disse friluftaktivitetene på fritida (ikke med skolen)?**

	Aldri	1 gang	2-5 ganger	Mer enn 5 ganger
Tursykling (utenfor offentlig vei/tettbygd strøk)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fottur i skog og mark	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fottur i fjellet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Klatring/buldring ute	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Padling i kano/kajakk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tur med overnatting i telt eller under åpen himmel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ferskvannsfiske	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sjøfiske	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jakt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bading (i sjø/elv/på kysten)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Plukking av sopp eller bær (bær- eller sopptur)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skøyter utendørs (på tur eller som leik)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skitur og/eller "frikjøring" på snø	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ridetur i skog og mark eller fjell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Svar ja eller nei ut fra om du har gjort dette i kroppsøving siden du begynte på videregående.**

	Ja	Nei
Jeg har planlagt og gjennomført oppvarming	<input type="radio"/>	<input type="radio"/>
Jeg har planlagt og gjennomført oppvarming flere ganger	<input type="radio"/>	<input type="radio"/>
Jeg har vært med på å planlegge og gjennomføre tur ut i naturen	<input type="radio"/>	<input type="radio"/>
Jeg har øvd på førstehjelp som er vanlig ved idrettsskader	<input type="radio"/>	<input type="radio"/>
Jeg har vært ute og brukt naturen som matkilde	<input type="radio"/>	<input type="radio"/>
Jeg har øvd på å kunne demonstrere og bruke gode arbeidsteknikker og arbeidsstillinger	<input type="radio"/>	<input type="radio"/>
Jeg har fylt ut skriftlig egenvurdering i kroppsøving(på papir eller data)	<input type="radio"/>	<input type="radio"/>
Jeg har hatt samtale med kroppsøvingslærer der jeg fikk tilbakemelding	<input type="radio"/>	<input type="radio"/>
Jeg har vært med og vurdert medelever i kroppsøving	<input type="radio"/>	<input type="radio"/>
Jeg har brukt reknekunnskaper (matematikk) i arbeid med kroppsøving	<input type="radio"/>	<input type="radio"/>
Jeg har brukt digitale verktøy (feks PC, pulsklokke) i arbeid med kroppsøving	<input type="radio"/>	<input type="radio"/>
Jeg har fått vurderinger som har ført til at jeg har skjerpet meg faglig	<input type="radio"/>	<input type="radio"/>



**Gjør en vurdering av hvor godt du klarer deg i/mestrer aktivitetene nedenfor.**

	1 Kan ikke/aldri prøvd	2	3	4	5 Mestrer godt
Dans du selv er med på å lage	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Volleyball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Basketball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Innebandy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Håndball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fotball	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Egentrening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tester (av ulike slag)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Friluftsliv (f.eks. tur i skogen, matlaging ute)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Styrketrening med vekter/treningsapparater	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Friidrett (f. eks. lenghe hopp, høyde, baneløp, kulestøt)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Orientering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Turn/akrobatikk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Norsk folkedans (f. eks. reinlender, polka, halling)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skigåing – langrenn, tur eller skileik	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ski, snøbrett, eller liknende i alpinbakke	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Skøyter/isspill	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Noe du til slutt vil si om kroppøving?**

Ordet fritt

**Tusen takk for at du svarte på undersøkelsen!**



Vedlegg 3:

Samtykkeerklæring for fokusgruppeintervju





## **Forespørsel om deltaking i gruppeintervju i forbindelse med forskningsprosjektet "Vurdering for læring i kroppsøving?"**

I arbeidet med min doktorgrad studerer jeg vurderingspraksis i kroppsøving ved flere videregående skoler. Skolen du arbeider ved er en av skolene som har sagt seg villig til å være med i prosjektet, og dere har tidligere fått informasjon om at prosjektet blant annet består av en nettbasert spørreundersøkelse blant elevene og et samtalebasert gruppeintervju med lærere som underviser i kroppsøving.

I intervjuet er målet å få fram meninger, intensjoner og praksis i forhold til vurdering i kroppsøving ved deres skole. Jeg er interessert i hva dere gjør i kroppsøving, da spesielt i forhold til vurdering, men også mer generelle tanker om kroppsøvingsfaget. Videre hva dere tenker om kunnskapsløftet, om dere har diskutert forhold rundt vurdering for læring, om det er spesielle utfordringer med enkelte elevgrupper, etc. Det brede bilde av vurderingspraksis som kommer fram gjennom gruppeintervjuet vil jeg i prosjektet drøfte mot resultater fra undersøkelsen blant elevene og resultater fra andre skoler.

Under intervjuet vil jeg vil bruke lydopptaker og gjøre enkelte notater mens vi snakker sammen. Intervjuet er planlagt å vare fra en til to timer.

Det er frivillig å være med og du har mulighet til å trekke deg når som helst underveis, uten å måtte begrunne dette nærmere. Dersom du trekker deg vil alle innsamlede data om deg slettes. Generelt gjelder at alle opplysninger vil bli behandlet konfidensielt, og både skole og enkeltpersoner anonymiseres ved formidling/rapportering av resultater. Lydopptak fra intervju slettes når prosjektet er ferdig, senest i 2014. På grunn av eventuelle oppfølgende studier, foretas fullstendig anonymisering der også navn på skoler slettes, senest i 2020. Prosjektet er godkjent av Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste A/S.

Med vennlig hilsen  
Petter Erik Leirhaug  
Stipendiat ved Norges idrettshøgskole  
Mobil: 47 76 03 70

### **Samtykkeerklæring:**

Jeg har mottatt muntlig og skriftlig informasjon om forskningsprosjektet *Vurdering for læring i kroppsøving?* og er villig til å delta i gruppeintervju

Dato ..... Signatur .....



Vedlegg 4:

Temaguide for fokusgruppeintervju





## Temaguide fokusgruppeintervju

Før oppstart: Informasjon om prosjektet. Samtykkeerklæring.  
Bakgrunnsinformasjon: kjønn/alder/utdanning/arbeidserfaring.

Kroppsøving ved skolen?  
Status/støtte fra ledelsen?

(Styrt) Arbeid med læreplan etter LK06?  
Hva – hvordan – hvem  
(Organisering? Diskuterer mye sammen, om kriterier, planer etc?)

Rammefaktorer:  
Hvor foregår undervisningen?  
Utstyr/Garderobefasiliteter?  
Spesielle elevforhold? Annet?

Mening og mål med faget i skolen?  
Innhold? (Kjennskap til læreplanen?)  
Hovedområdene – god fordeling?  
    Idrett og dans  
    Friluftsliv  
    Trening og livsstil  
Periodeplaner?

### **Hvordan vil dere beskrive vurdering i kroppsøving ved denne skolen?**

**Hva legger dere vekt på ved vurdering?  
Hvordan formidles dette til elevene?**

Diskuterer dere ofte vurdering i kroppsøving?

Spesielle utfordringer ved vurdering i kroppsøving?

Har dere utarbeidet vurderingskriterier/kjennetegn på måloppnåelse?  
(Hvordan er disse utarbeidet? Har alle vært med?)

Tror dere elevene vet hva de blir vurdert i forhold til? Vet de f.eks. hva som skal til for å oppnå de ulike karakterene?

Er det vanskelig å sette (rettferdige) karakterer i kroppsøving?

Forhold mellom sluttvurdering og undervisvurdering?  
(Hva med innsats? og elevenes forutsetninger?)

Får elevene tilbakemeldinger i kroppsøvingstimen? (tilbakemelding på hva; læring?)

Hva med teoriprøver? (Hvordan bruker dere prøvene?)

Hva med tester? (Følg opp; hvilke? Beskriv.)

Hva med egentrening i forhold til vurdering?

Har dere (fag)samtale med elevene? (hvordan – hvor ofte?)

Egenvurdering? (bruker dere det?)

Medelevvurdering / Hverandrevurdering?

Er elevene med på å utforme kriterier/formulere egne mål?

Bruker dere læreplanen aktivt i vurderingsarbeidet? (med elevene?)

**Hvis jeg sier vurdering FOR læring i kroppsøving, hva tenker dere da?**

(Gjerne eksempler; få dem til å utdype/beskrive)

På hvilke måter gir dere feed-back til elevene i forhold til kompetansemål?

I hvor stor eller liten grad vil dere si at dere greier å gi slik tilbakemelding til elevene?

Vil dere si at ny forskrift for vurdering og læreplanen i faget har ført til noen endringer av innhold og arbeidsmåter i kroppsøving?

Fagets fremtid?

Annet dere vil si eller tilføye om vurdering i kroppsøving?

Om kroppsøving?

Vedlegg 5:

Godkjenning av prosjektet fra NSD (Norsk samfunnsvitenskapelig datatjeneste – personvernombudet for forskning)





Harald Hårfagres gate 29  
N-5007 Bergen  
Norway  
Tel: +47-55 58 21 17  
Fax: +47-55 58 96 50  
nsd@nsd.uib.no  
www.nsd.uib.no  
Org.nr. 985 321 884

Petter Erik Leirhaug  
Seksjon for kroppsøving og pedagogikk  
Norges idrettshøgskole  
Postboks 4014 Ullevål stadion  
0806 OSLO

Vår dato: 16.02.2011

Vår ref: 26065 / 3 / LMR

Deres dato:

Deres ref:

## TILRÅDING AV BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 17.01.2011. Meldingen gjelder prosjektet:

26065

*Vurdering for læring i kroppsøving? En undersøkelse av vurderingspraksis og elevenes oppfatninger av vurdering i kroppsøving ved åtte videregående skoler, og en toårig intervensjon med mål om å bedre vurderingspraksis ved fire av de videregående skolene Norges idrettshøgskole, ved institusjonens overste leder  
Petter Erik Leirhaug*

Behandlingsansvarlig  
Daglig ansvarlig

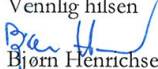
Personvernombudet har vurdert prosjektet, og finner at behandlingen av personopplysninger vil være regulert av § 7-27 i personopplysningsforskriften. Personvernombudet tilrår at prosjektet gjennomføres.

Personvernombudets tilråding forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, eventuelle kommentarer samt personopplysningsloven/-helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, [http://www.nsd.uib.no/personvern/forsk\\_stud/skjema.html](http://www.nsd.uib.no/personvern/forsk_stud/skjema.html). Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://www.nsd.uib.no/personvern/prosjektoversikt.jsp>.

Personvernombudet vil ved prosjektets avslutning, 31.12.2014, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen  
  
Bjørn Henriksen

  
Linn-Merethe Rød

Kontaktperson: Linn-Merethe Rød tlf: 55 58 89 11  
Vedlegg: Prosjektvurdering



Vedlegg 6:

Innhold i prosjektets databehandleravtale med Websurvey hos NSD (Norsk samfunnsvitenskapelig datatjeneste – personvernombudet for forskning)





**Avtale om behandling av personopplysninger mellom Norges idrettshøgskole (behandleransvarlig) og Norsk samfunnsvitenskaplig datatjeneste (databehandler)**  
(jfr. Personopplysningslovens §15.)

**Prosjektittel: Vurdering for læring i kroppsøving?**

Prosjektleder/daglig ansvarlig: Petter Erik Leirhaug  
Surveynummer NSD: 139  
Prosjektnummer NSD:26065

1. I forbindelse med innsamling av forskningsdata i ovennevnte prosjekt skal databehandler behandle personopplysninger i henhold til kvittering fra Personvernombudet for forskning ved NSD (jfr. personopplysningsloven §31 eller personopplysningsforskriften §7-27), fra tilsvarende institusjon eller fra Datatilsynet (jfr. personopplysningsloven §33).
2. Databehandler kan bare behandle personopplysninger fra ovennevnte prosjekt i forhold til avtalt forskningsformål. Direkte identifiserende opplysninger som navn og epostadresser lagres i en database i kryptert form og brukes kun til elektronisk utsendelse av invitasjoner eller påminnelser i forbindelse med innsamling av data for prosjektet. Disse opplysningene slettes så snart det er avklart at det ikke vil bli sendt ut nye påminnelser, og senest ved utløp av avtalen.
3. Datamaterialet skal behandles i henhold til det som er spesifisert i kvittering fra personvernombudet eller i konsesjon fra Datatilsynet. Dersom personopplysninger skal oppbevares utover 01.10.2014, må behandlingsansvarlig forelegge godkjenning for dette fra personvernombudet eller fra Datatilsynet.
4. Databehandler er pliktig til å gjennomføre sikkerhetstiltak som følger av personopplysningslovens §13. Databehandler skal videre sørge for at dokumentasjon av informasjonssystemet og sikkerhetstiltakene er tilgjengelig for behandlingsansvarlig, Datatilsynet og Personvernemnda.
5. Databehandler kan bare utlevere personopplysninger til andre enn behandlingsansvarlig etter særskilt avtale med behandlingsansvarlig og bare når det foreligger gyldig hjemmelsgrunnlag for slik utlevering samt tilråding fra Personvernombudet for forskning ved NSD eller fra tilsvarende institusjon, eller konsesjon fra Datatilsynet.
6. Denne avtalen utløper 01.10.2014. Senest 14 dager tidligere (17.09.2014) slettes alle personopplysninger og innsamlete data fra NSDs databaser, og data overføres til behandlingsansvarlig. Data kan oppbevares offline hos NSD til avtalens utløp.
7. Data som overføres til behandlingsansvarlig skal ikke kunne kobles til personopplysningene som er brukt for å samle inn data med mindre det er eksplisitt avtalt og tillatelse foreligger.





