

Julie Marie Isager
Syddansk Universitet

DOI: <http://dx.doi.org/10.5617/adno.8204>

Konstrukt, generalisering og held ved mundtlige eksaminer i gymnasiet

Sammendrag

Danske gymnasieelever bliver – ligesom norske – eksamineret ved mundtlig præsentation og samtale om et udtrukket emne af deres lærer og en censor fra et andet gymnasium. Disse eksaminer undersøges ikke systematisk, og der forskes ikke i dem. Det samlede eksamensresultat er high-stakes og får konsekvenser for elevens fremtidige uddannelsesmuligheder. Artiklen undersøger centrale danske styredokumenter, der regulerer mundtlige danske eksaminer i gymnasiet ved at lægge et teoretisk blik hentet fra international anerkendt assessmentteori: Hvad intenderer eksamen at måle? Assessmentteoretiske begreber defineret af Standards for Educational and Psychological testing og Michael T. Kanes argumentationsbaserede tilgang til validitet bruges til at søge efter et *konstrukt* og slutninger i eksamenssystemet. Data er lovdokumententer samt læreplan og vejledning i historiefaget som eksempel i det almene gymnasium med et almindelig og studieforbereende formål. Dermed er artiklen et tillæg til Acta Didactica Norges temanummer vol. 12 nr. 4 om Test og eksamen i Sverige og Norge. Studiet finder, at konstruktet er vanskeligt at identificere. Antagelser om mundtlige eksaminer er ikke ekspliciterede i dokumenterne, og slutninger må analyseres frem. Eleven eksamineres i et dobbelt udtræk- eller samplingssystem: Eksamensfaget udtrækkes og emnet inden for disciplinen udtrækkes. Der etableres dermed en antagelse om vidtgående generaliseringsmulighed af testresultatet på tværs af fag. Det diskuteres om assessmentteori egner sig til diskussion af traditionsrige danske mundtlige eksaminer.

Nøkkelord: mundtlig eksamen, assessment, konstrukt, gymnasiet, Danmark

Construct, Generalization and Luck in Danish Oral Examinations in Upper Secondary School

Abstract

Danish and Norwegian upper secondary school students are assessed on their presentation and discussion of a sampled topic by their teacher (the examiner) and a teacher from another school (external examiner). These oral exams are not investigated systematically even though they are high stakes tests with considerable consequences for the students' potential choice of further education. The article investigates the *construct* of the oral exam: What are oral exams intended to assess? Data are key law documents, the curriculum and the instructions given for the subject of History as an example. The documents are read with assessment theory as defined in the Standards

for Educational and Psychological Testing and Michael T. Kane's argument-based approach to validity. The study finds that the construct is difficult to identify. No explicit arguments or inferences are found in the documents read, and inferences must be interpreted between the lines. Each student is assessed in a double sampling system: The subject is sampled, and the theme of the exam is sampled within the subject. The article shows a system built with inferences of extensive generalization from one score interpretation to potential other performances across the curricula. It is discussed whether assessment theory is adequate for the Danish educational traditions.

Keywords: oral exam, assessment, construct, upper secondary school, Denmark

Introduktion

”Velkommen, kom indenfor og træk et spørgsmål”, siger eksaminator smilende til eksaminanden, der ser nervøs ud ude på gangen. I lokalet ligger en stribe små sedler med bagsiden op klar til træk. Censor står ved bordet med den traditionsrige grønne dug.

Mundtlige eksaminer har været en central testmetode i danske eksaminer i gymnasiet siden 1848 (Haue et al., 1998, s. 33–34). Den bliver specificeret gennem årene, men ændrer ikke navn eller status. Fra 1964 prøves elever ikke i alle fag, men kun i fag der udtrækkes (Undervisningsministeriet, 1964), og nutidige eksaminer har dermed historisk ballast, svarende til norske mundtlige eksaminer (Uddannelsesdirektoratet, 2019a). Eksaminerne er high-stakes¹ for de 18–20-årige elever, da optagelse til danske videregående uddannelser administreres efter studentereksamens karaktergennemsnit². Optagelsessystemet til danske universiteter diskuteres³, og der har været enkelte rapporter med forslag til revision af eksamensformer og censorsystemet (Danmarks Evalueringsinstitut, 2003; Undervisningsministeriet, 2008), men hverken danske studentereksaminer eller mundtlige eksaminer diskuteres offentligt, selvom stort set alle voksne har erfaringer med testmetoden, fordi den også findes i folkeskolen. Litteratursøgning efter dansk forskning i mundtlige studentereksaminer giver kun få hits.⁴ Tofteskov og Leth Andersen (2016) introducerer til danske eksamensformer, og eksamensformers validitet er undersøgt sparsomt i universitetskontekst (O'Neill et al., 2019; Skov, 2013). I Norge debatteres eksaminer offentligt, men der forskes kun begrænset, beskriver Uddannelsesdirektoratets ”Eksamensgruppa”, der giver et forskningsoverblik over et bredt antal spørgsmål angående slutvurdering (Uddannelsesdirektoratet, 2019a). Acta Didactica Norges temanummer (vol. 12,

¹ En term for, at testresultatet får håndgribelige konsekvenser, se evt. Standards (2014) s. 188–189.

² Studiepladser til videregående studier fordeles via kvote 1 (gennemsnit af eksamens- og årskarakterer fra gymnasiet) og kvote 2 (gennemsnit af eksamens- og årskarakterer fra gymnasiet rækker næsten, og det videregående uddannelsessted kan stille supplerende krav i forbindelse med optaget). Der er få studiepladser i kvote 2.

³ Det danske Uddannelses- og Forskningsministeriet skriver om fokus på et nyt optagelsessystem i foråret 2019 på <https://ufm.dk/uddannelse/indsatsomrader/bedre-optagelsessystem>.

⁴ Se bilag med søgestreng og databaser.

nr. 4, 2018) har svenske og norske studier af prøver og eksamen i skolen, men ingen danske studier.

Den spredte internationale forskning i mundtlig eksamen beskæftiger sig kun sjældent med, hvad der måles (se dog Borger, 2018; Bøhn, 2016; Dobson, 2008; Palmér, 2010). Hovedsagelig tematiserer forskningen specifikke elev- og lærer-erfaringer eller anskuer mundtlig eksamen som en alternativ eksamensform (Huxham et al., 2012; Joughin, 2003; Kalthoff, 2013; Kvifte, 2011; Maugesten, 2011; Pearce & Lee, 2009; Rolin, 2013).

Internationalt forskes og udvikles test og eksamen under termen *assessment*. Omfattende amerikansk *assessment*-forskning sammenfatter kvalitetstandarder for eksaminer i konsensusdokumentet *Standards for Educational and Psychological Testing* (forkortet til *Standards*) forfattet af organisationerne American Educational Research Association, American Psychological Association og National Council on Measurement in Education (Standards, 2014). Her er udgangspunktet begrebet *the construct*, der i Sverige og Norge omtales som *konstruktet* (fx Erickson & Gustafsson, 2014; Uddannelsesdirektoratet, 2019a).

Denne artikels forskningsspørgsmål undersøger, hvad der sker, hvis danske mundtlige eksaminer i gymnasiet læses med *assessment*-teori: Hvad er danske mundtlige studentereksamener konstrukt? Samt om *assessment*-teori er relevant for et studie af en eksamensform i en dansk didaktisk tradition? Artiklen bidrager til nordisk forskning i mundtlige eksaminer med et teoretisk fremmed blik fra international *assessment*-teori på de officielle rammer for en dansk traditionel eksamensform med substantielle konsekvenser for eleverne. Først præsenteres *assessment*-teori, og derefter fremlægges den analytiske operationalisering i metodeafsnittet, der også præciserer de læste styredokumenter. Mundtlige eksaminers konstrukt søges fremanalyseret i en såkaldt konstruktaspektanalyse, der gennemføres på uddannelsesniveau og på fagniveau. Grundlæggende slutninger i eksamenssystemet analyseres med Kanes argumentationsbaserede tilgang til validitet. Herefter diskuteres analysefund og teoriens forklaringskraft i en dansk uddannelsestradition.

Assessmentteori

Den amerikanske standard baserer sig på refleksioner over *assessment* siden 50'erne (Cronbach & Meehl, 1955; Kane, 2013; Messick, 1989). American Psychological Association (APA) opregner 66 tidsskrifter, der beskæftiger sig med *assessment* kvalitativt og kvantitativt i uddannelse og psykologi⁵, og der er centerdannelser i USA og Europa, der løbende undersøger validiteten af test og eksamination i uddannelsessystemet⁶. *Konstruktet* står centralt og defineres:

⁵ <https://www.apadivisions.org/division-5/resources/journals/index> , læst 28.maj 2019.

⁶ Fx Oslo Universitets Centre for Educational Measurement, Dublin City University Centre for Assessment Research, Policy and Practice in Education, Helsinki Universitets Centre of Educational

The term *construct* is used in the *Standards* to refer to the concept or characteristic that a test is designed to measure. Rarely, if ever, is there a single possible meaning that can be attached to a test score or a pattern of test responses. Thus, it is always incumbent on test developers and users to specify the construct interpretation that will be made on the basis of the core or response pattern. (Standards, 2014 s. 11)

Et eksempel er ”mathematics achievement” (Standards, 2014, s. 11), der som alle andre konstruktter må fortolkes, præciseres og udfordres. Konstruktet er påvirket af mangefacetterede logikker og processer – en konstruktion, som termen konstrukt så tydeligt indikerer. Det eksamen udprøver, eksisterer ikke som afgrænset endegyldig enhed, men praktiseres socialt under indflydelse af mange parametre (Messick, 1989). De sociale praksisser konstruktet findes i, må opdeles for at kunne identificeres systematisk, og det gøres ved at pege på aspekter, som indikerer konstruktet (Messick, 1989, s. 17). Konstruktet fungerer som referencepunkt for validering.

Validering defineres som processer hvor teoretiske og empiriske belæg hentes som dokumentation for, at elevens eksamensresultat og dets brug er gyldigt i forhold til formålet (Standards, 2014, s. 11; Utdanningsdirektoratet, 2019a, 4.1). Validering foregår som successive undersøgelsesprocesser af praksisser og slutninger, der påvirkes af normer, og er dermed gradsspørgsmål. Eksamination kan ikke bedømmes universelt enten valid eller ikke-valid, eftersom ”[t]ests do not have reliabilities and validities, only test responses do” (Messick, 1989, s. 14). Et klart defineret konstrukt er et udgangspunkt for valide eksamensresultater, men mange aspekter af eksamensforløb må undersøges for at validere resultatet:

Validation involves careful attention to possible distortions in meaning arising from inadequate representation of the construct and also to aspects of measurement, such as test format, administration conditions, or language level, that may materially limit or qualify the interpretation of test scores for various groups of test takers. (Standards, 2014, s. 13)

Historisk har assessmentteori flyttet sig fra fokus på testens indhold (content validity) og sæt af referencekriterier (criterion validity) til de to i forening: om testen måler det, den hævder (construct validity). Dermed fokuserer validering på de sociale processer, som testen optræder i. Spørgsmålet flytter sig fra validitet til validering, da Messick i 1989 samler tidligere validitetstyper i construct validity for at fokusere på spørgsmål om eksamens retfærdighed for den enkelte elev (Messick, 1989, 1995, s. 741). I stedet for isoleret at undersøge testens enheder (i multiple choice-formater kaldet *items*), blev en bred tilgang til validering at fokusere på konsekvenserne af testen for de involverede parter: elever, forældre, lærere og administration, uanset testmetoden. Idealet er at sikre, at assessment ikke kolliderer med sociale værdier om retfærdighed med formålet, at ”social consequences should not be attributed to any source of test invalidity” (Messick,

1995, s. 748), og at validering bliver en kontinuerlig proces mellem mange interessenter. Denne teori om assessment dækker alle typer af testmetoder og ikke kun multiple choice-formater (Standards, 2014).

Messick beskrev i 1989 to *trusler mod validitet*, der har dannet præcedens i en grad, så Standards opfører termerne uden at henvise til Messick (Standards, 2014, s. 12 og 217; Messick, 1989, 1995). Kane henviser gentagne gange til dem (Crooks et al., 1996; Kane, 1992, 2006, 2013):

- *Construct underrepresentation*: Prøven tester for snævert i forhold til det intendede konstrukt.
- *Construct-irrelevant variance*: Prøven tester bredere end det intendede konstrukt; der udprøves flere parametre end ekspliciteret.

Det er ikke entydigt at afgøre konstruktrelevans eller -irrelevans. Messick eksemplificerer med, at kommunikative evner ved en matematiktest vil kunne vurderes som både relevant og irrelevant alt efter konstruktfortolkningen (Messick, 1995, s. 743), hvilket accepteres i Standards (2014, s. 12). Truslerne mod konstruktvaliditeten gælder både for store nationale assessmentprogrammer og for mere lokal bedømmelse.

Kanes argumentbaserede tilgang til validitet operationaliserer Messicks krav til det nye udvidede konsekvensperspektiv i to faser: 1. *Interpretation/Use Argument (IUA)* og 2. *Validity argument* (Kane, 1992, 2013, s. 8).

1. IUA ekspliciterer validitetsargumenterne i slutninger, der præciserer, hvordan testen rettelig undersøger det intendede konstrukt (Kane, 2013). Interpretation eller fortolkninger defineres hos Kane som: "[T]o explain the meaning of the score and, thereby, to make at least some of the implications of the score clear" (Kane, 1992, s. 527). Ved at acceptere både fortolkninger og praksisser som grundlag for IUA kan både skrivebords- og feltstudier bruges som datagrundlag. Fire slutningstyper er almindeligt forekommende: Scoring (at oversætte en observation til en bedømmelse), generalisering (at bedømmelsen af en observation kan generaliseres til andre observationer), extrapolation (at bedømmelsen kan overføres til en ikke-skolekontekstuel situation) og implications (hvordan bedømmelsen indgår i eller informerer handlinger) (Kane, 1992, s. 529, 2006, s. 24, 2013, s. 25). Generaliseringsantagelsen beskrives således:

The observations are treated as if they have been sampled from some universe of observations, involving different occasions, locations, and observers that could have served equally well; in generalizing over conditions of observation, one draws conclusions about the universe of possible observations on the basis of a limited sample of actual observations. The assumptions supporting such inferences are invariance laws stating that the conditions of observation involved in the measurement can be allowed to vary along certain dimensions without changing the outcomes. (Kane, 1992, s. 529)

Generalisering af et testresultat til andre kontekster står centralt i eksamensvalidering (Standards, 2014, s. 18). Kriterierne for IUA er 1. klarhed, at argumentationen for validiteten af eksamensresultatet er tilstrækkelig ekspliciteret, så

slutninger kan kritiseres og afprøves, 2. kohærens mellem argumentets dele, og 3. rimelighed, argumentationens antagelser anses for at være gyldige (Kane, 1992). Et argument kan kritiseres ud fra alle tre kriterier, men "weak assumptions, especially weak "hidden" assumptions, are typically the most serious problem" (Kane, 1992, s. 528). Standards opregner lignende kriterier for klarhed og løbende valideringsundersøgelser:

1.0 Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided. (Standards, 2014, s. 23)

Det er testdesigners opgave at klargøre testens intenderede konstrukt, målgruppe og brug – at der skal leveres et rationale og bevis for, at testdesignet er adækvat og tydeligt deklareret (Standards, 2014, 1.1, 1.2 og 1.3 s. 23).

2. I *Validation Argument* undersøger man systematisk, om slutningerne i IUA holder ved at udfordre argumenterne med forskellige typer af empiriske studier. Standards bruger udtrykket "validity evidence" for disse empiriske undersøgelser og skriver (2014, s. 21), at et "sound validity argument integrates various strands of evidence". Det er argument og slutninger der valideres med konstruktet som udgangspunkt (Kane, 2006).

De to faser adskilles primært af deres intention om at kortlægge og at undersøge holdbarheden af validitetsargumentationen. Inspirationen til den argumentationsbaserede validering er hentet fra bl.a. Toulmin (2003) og Perelman og Olbrechts-Tyteca (1969) – et praktisk argumenterende udgangspunkt med afsæt i retorik frem for formel logik (Kane, 1992, 2006, 2013). Det giver en fleksibel teoretisk ramme, der kan tilpasses til konteksten (Kane, 2013, s. 10). Der er ikke bestemte mønstre eller tjeklister for valideringen, og Kane henviser til Toulmins argumentmodel (Toulmin, 2003) som redskab for at eksplicite og kritisere påstande, belæg og hjemler for slutninger (Kane, 2013, s. 11–12). Kane eksemplificerer en påstand – "Scores on the test are generalizable across samples of items, scorers, and occasions." – og forslår generaliseringsstudier som belæg for eksamensvaliditet (Kane, 1992, s. 531). Eksamensvaliditet er normbaserede diskussioner, der ikke finder et endelig, entydigt svar.

Metode og empiri

Elevens studentereksamen afholdes i et antal fag, som udtrækkes blandt alle fag, ligesom den norske trekordning (Uddannelsesdirektoratet, 2019a). Undersøgelsen af konstruktet i mundtlige eksaminer i en dansk studentereksamen kan ikke lade analysen blive inden for rammerne af det enkelte fag, da den er underlagt overordnede formål, eksamensbestemmelser, tidsplaner og aktører.

Artiklens data er fire styredokumenter, som tilsammen er de vigtigste bestemmelser for elevernes eksaminer: to på uddannelsesniveau og to på fagniveau. På

et overordnet uddannelsesniveau præciserer *Lov om de gymnasiale uddannelser* (2016), hvad en studentereksamen er og består af. Kapitel 1 om uddannelsernes formål og kapitel 6 om eksamen er relevante for undersøgelsen af konstruktet på uddannelsesniveau. Loven er politisk vedtaget, og politikerne er ansvarlige for det danske eksamensdesign, inklusive testmetodernes udformning. Lovens forarbejder kunne eksplicitere antagelser under eksamenssystemet, men det har vist sig, at forarbejderne ikke indeholder overvejelser om eksamens indretning hverken på uddannelses- eller fagniveau ud over et enkelt ønske om, at ”prøveformerne udvikles, så de rent faktisk afspejler, hvad elever skal kunne i dag” (Betænkning over Forslag til lov om de gymnasiale uddannelser, 2016). Loven henviser for eksamensbestemmelser til *Bekendtgørelse om prøver og eksamen i de almene og studieforberevende ungdoms- og voksenuddannelser* (Eksamensbekendtgørelsen, 2016), et dokument på i alt 13 sider, der, som titlen antyder, dækker alle de gymnasiale retninger og voksenuddannelserne med hver deres profiler og formål. Elevers konkrete eksaminer på fagniveauet er beskrevet i *læreplaner* i bilag til loven (Undervisningsministeriet, 2017). Læreplanen til det enkelte fag følges af en *vejledning*, der ”præciserer, kommenterer, uddyber og giver anbefalinger vedrørende udvalgte dele af læreplanens tekst, men indfører ikke nye bindende krav.” (Undervisningsministeriet, 2018)⁷. På fagniveau er historiefaget udvalgt som eksempel, fordi det er et omfattende fag på højt niveau og er obligatorisk på alle linjer i det almene, studieforberevende gymnasium. Eksamensformen er ændret i seneste reform fra 24 timers forberedelse til 90 minutter.

Dokumentanalysen af mundtlige danske studentereksaminer tager udgangspunkt i to lag af analyse. Første analyse kalder jeg en konstruktaspektanalyse, og jeg har udarbejdet et tekstanalytisk instrument til at identificere og kortlægge aspekter af konstruktet med inspiration fra assessmentteori. Ved at læse konstruktaspekterne i sammenhæng dannes et grundlag for den anden analyse, der udgøres af Kanes Interpretation/Use Argument (IUA), hvor de slutninger som eksamen bygger på, fremlægges (Kane, 1992, 2006, 2013). Bidraget i den anden analytiske fase er at fortolke de slutninger, som lovtæksten må bygge på, og dermed kan de eksplicitte udsagn på uddannelsesniveau og på fagniveau fremstå sammenhængende.

Konstruktaspektanalysen er en hermeneutisk læsning designet med udgangspunkt i den internationale assessmentlitteratur. Her identificerer jeg aspekterne hver for sig i dokumenterne og læser dem derefter sammen for at skabe indsigt i, hvordan de interagerer og med- eller modarbejder hinanden. Termen *aspekt* er hentet fra Standards (2014, s. 11). Konstruktaspekterne udarbejdet til dette dokumentstudium fremgår af tabel 1.

⁷ Den analyserede vejledning fra marts 2018 kan rekvireres hos forfatteren. Den er ikke længere tilgængelig online, da den er erstattet af en ny.

Tabel 1. Konstruktaspektspecifikation

Konstrukt-aspekt	Konstruktaspektspecifikation	Henvisninger
Formål med eksamen	Beskriver den intention, der ligger bag testen, den funktion den skal opfylde for de implicerede parter og for samfundet. Knyttet til anvendelse (Use) og konsekvens. ”Assessment for what?” hos Gipps.	Standards s. 14 og 184–188; Gipps 2012 s. 2–4
Kriterier for eksamen	Beskriver de kriterier, som eksamenspræstationen skal leve op til. Historisk knyttet til criterion validity, test criterion og criterion domain.	Standards s. 17, 218; Kane 2006 s. 18
Testens indhold	Beskriver eksamens indhold med reference til litteratur, temaer og lignende. Knyttet til test content og content domain.	Standards s. 14, 218; Kane 2006 s. 19
Test-metode	Eksamensformens præcise rammer med angivelse af tid, materialeomfang og bekendtgørelse af resultatet.	Standards s. 224
Aktører	Involverede parter med funktioner i forbindelse med eksamens gennemførelse. (Standards: test publisher, raters, test developer etc.)	Standards s. 220–224

IUA-argumentanalysen er foretaget på baggrund af tekstfortolkningen i konstruktaspektanalysen og har anvendt Toulmins argumentmodel som analyse-redskab. Uddannelsesniveaets lov og bekendtgørelse er læst simultant, derefter er fagniveaets læreplan læst og sidst vejledningen.

Analyse

Analyseresultaterne er fremstillet i to dele: Først konstruktaspektanalysen, der identificerer aspekter af konstruktet i styredokumenterne på uddannelsesniveau og fagniveau. Som eksemplificerende for, hvilke indsigter det kan give at foretage konstruktaspektanalyse inden formuleringer af IUA, præsenterer jeg eksempler på læsninger af kombinationer af konstruktaspekter. Derefter følger analyse-resultater fra analyse med Kanes argumentationsbaserede framework Interpretation/Use Argument.

Konstruktaspektanalyse på uddannelsesniveau

Konstruktaspekterne på uddannelsesniveaet identificeres i loven og bekendtgørelsen som anført i tabel 2.

Tabel 2. Konstruktaspektanalyse på uddannelsesniveau: Lov om de gymnasiale uddannelser 2016 (Lov) og Eksamenbekendtgørelsen (Eks.bek.).

Konstrukt-aspekt	Konstruktaspektspecifikation på uddannelsesniveau	Henvisninger
Formål med eksamen	1. Certificering af den enkelte elev: ” [A]t dokumentere, i hvilken grad eksaminanden opfylder de mål og krav, der er fastsat for faget og uddannelsen.”	Eks.bek. §1
Kriterier for eksamen	1. Uddannelsens formål: At forberede elever til videregående uddannelse, ”herunder at de tilegner sig almindelig viden og kompetencer gennem uddannelsens kombination af faglig bredde og dybde og gennem samspil mellem fagene”	Lov §1

Testens indhold	Ikke ekspliciteret	
Test-metode	<ol style="list-style-type: none"> 1. Der samples 10 eksamensfag for individuel elev. Obligatorisk er kun mundtlig eksamen i studieretningsprojekt 2. Eksamen ”tilrettelægges med henblik på at dokumentere graden af målopnåelse i forhold til væsentlige mål og krav”, og ”Prøveformerne skal samlet tilgodese uddannelsens formål” 3. Årskarakterer og eksamenskarakterer gives efter matematiske vægte knyttet til fagets niveau (A, B, C) 4. Testmetoden specificeres mht. materiale, forløb og præsentationsform i fagene⁸ 5. Opgaverne ved mundtlige eksaminer fordeles ved lodtrækning med mindre andet er angivet 6. Bedømmelsen meddeles straks eksaminanden 	<ol style="list-style-type: none"> 1. Lov §45 2. Eks.bek. §2 3. Lov §49, stk. 2; Eks.bek. §40 4. Eks.bek. §2 5. Lov §12 6. Eks.bek. §33
Aktører	<ol style="list-style-type: none"> 1. Ministeriet samler fag individuelt pr elev 2. ”Institutionen” udarbejder testmaterialet 3. Eksaminator er som hovedregel elevens lærer 4. En institutionsekstern censor⁹ skal ”medvirke til og påse, at eksaminanderne får en ensartet og retfærdig behandling, og at deres præstationer får en pålidelig bedømmelse” 5. Eksaminanden, som har rettigheder og pligter for at kunne aflægge eksamen 	<ol style="list-style-type: none"> 1. Lov §45, stk. 2 2. Eks.bek. §2, stk. 2 3. Eks.bek. §29, stk. 2 4. Eks.bek. §29 5. Eks.bek. §13–20

Uddybning af konstruktaspekterne *Formål og Kriterier for eksamen på uddannelsesniveau*

Formålet med studentereksamen, hvor mundtlige eksaminer indgår, er enkelt og kort beskrevet i eksamensbekendtgørelsen:

§ 1. Formålet med prøver og eksamener, som er reguleret i bekendtgørelsen, er at dokumentere, i hvilken grad eksaminanden opfylder de mål og krav, der er fastsat for faget og uddannelsen. (Eksamensbekendtgørelsen, 2016)

Der er et klart individuelt summativt certificeringsformål med eksamen (Harlen, 2007, s. 17). Andre mulige formål kunne være at støtte læringsforløb, at informere elever og forældre om elevens udbytte af gymnasieuddannelsen, at undersøge skolens undervisning eller at monitorere gymnasieskolen (Erickson & Gustafsson, 2014; Gipps, 1994, s. 3; Harlen, 2007, s. 16; Tveit & Olsen, 2018). Andre formål kan muligvis findes i aktørernes fortolkning og *use* af eksaminationerne. Formålsformuleringen er dobbelt ved at sidestille eksamens mål for faget *og* for uddannelsen, en relation der ikke er nærmere uddybet, hvilket lader den være åben for vid fortolkning. Læses Lov om de gymnasiale uddannelser fra 2016, er der ikke angivet *mål* for uddannelsen som

⁸ De mundtlige eksamensformer varierer med forberedelsestider fra ingen, 24, 60, 90 minutter til 24 timer. Materialet kan være selvproduceret, kendt, ukendt eller en kombination, og eksaminationstiden svinger fra 24 til 30 minutter med forskellige formuleringer af forventningerne til elevfremlæggelser og/eller dialog.

⁹ En lærer fra en anden skole rejser efter ministeriets plan for at eksaminere de enkelte elever, hvilket udgør den danske model for validering af slutningstypen Scoring hos Kane.

formålsbeskrivelsen i eksamensbekendtgørelsen fordrer, men der er angivet et *formål* for uddannelsen, der lyder således:

§ 1. Formålet med uddannelserne omfattet af denne lov er at forberede eleverne til videregående uddannelse, herunder at de tilegner sig almindelig viden og kompetencer gennem uddannelsens kombination af faglig bredde og dybde og gennem samspillet mellem fagene.

Stk. 2. Eleverne skal gennem uddannelsens faglige og pædagogiske progression udvikle faglig indsigt og studiekompetence. De skal opnå fortrolighed med at anvende forskellige arbejdsformer og opnå evne til at fungere i et studiemiljø, hvor kravene til selvstændighed, samarbejde og sans for at opsøge viden er centrale.

Stk. 3. Uddannelserne skal have et dannelsesperspektiv med vægt på elevernes udvikling af personlig myndighed. Eleverne skal derfor lære at forholde sig reflekterende og ansvarligt til deres omverden: medmennesker, natur og samfund samt til deres udvikling. Uddannelserne skal tillige udvikle elevernes kreative og innovative evner og kritiske sans.

Stk. 4. Uddannelserne og institutionskulturen som helhed skal forberede eleverne til medbestemmelse, medansvar, rettigheder og pligter i et samfund med frihed og folkestyre. Undervisningen og hele institutionens dagligliv må derfor bygge på åndsfrihed, ligeværd og demokrati og styrke elevernes kendskab til og respekt for grundlæggende friheds- og menneskerettigheder, herunder ligestilling mellem kønnene. Eleverne skal derigennem opnå forudsætninger for aktiv medvirken i et demokratisk samfund og forståelse for mulighederne for individuelt og i fællesskab at bidrage til udvikling og forandring samt forståelse af såvel det nære som det europæiske og det globale perspektiv.

(Lov om de gymnasiale uddannelser, 2016)

Hvis formålet kan forstås som konstrukt for den samlede studentereksamen, må det betegnes som bredt og upræcist. Det omfatter mål for både elevernes udvikling undervejs i uddannelsen, deres indstilling til denne udvikling, deres myndighed, fordybelse, faglige bredde, deres interaktioner og kulturelle værdier som demokrati, kønnenes ligestilling og globalt udsyn. Disse formål er vanskelige at udprøve, vanskelige for eksaminanderne at demonstrere i de faktiske mundtlige eksaminer i de fag, der samples til den enkelte elev. Eksamensbekendtgørelsen understøtter læsningen af, at formålet med uddannelsen er et eksamenskonstrukt, eftersom den præciserer: ”§ 11. Prøveformerne skal samlet tilgodese uddannelsens *formål*” (min fremhævelse). Hvordan det sker i praksis, eller hvordan sammenhængen mellem eksamensformerne og formålet tænkes at tilgodese hinanden, ekspliciteres ikke.

Konstruktaspektanalyse på fagniveau

Konstruktaspekterne på fagniveau (historiefaget) identificeres i læreplan og vejledning som anført i tabel 3.

Tabel 3. Konstruktaspektanalyse af historiefaget STX: læreplan og vejledning (vejl.) (Undervisningsministeriet, 2017, 2018).

Konstrukt-aspekt	Konstruktaspektspecifikation på fagniveau – Historie A STX	Henvisninger
Formål med eksamen	1. Bedømme ”i hvilken grad eksaminandens præsentation opfylder de faglige mål”	1. Læreplan 4.3
Kriterier for eksamen	1. 11 læringsmål formuleret med udgangspunkt i taksonomiske verber 2. <i>Eksempel</i> på hvordan karakterskalaens trin <i>kan</i> afspejle læringsmål 3. ”Det vil normalt ikke være muligt for eleven/kursisten at demonstrere en udfoldet opfyldelse af alle faglige mål, men det tilstræbes at der gennem materialesæt og eksamination åbnes for målopfyldelse af et betydeligt antal heraf.” 4. Præsentationen indeholder ”som minimum problemstilling(er), konklusion(er) og evt. anvendte metoder”	1. Læreplan 2.1 2. Vejl. s. 32 3. Vejl. s. 31 4. Vejl. s. 31
Testens indhold	1. Kernestof er beskrevet med overskrifter for forløb, der danner overskrifter for materialesæt, fx Ideologierne kamp.	1. Læreplan 2.2
Testmetode	1. 3–5 siders ukendt forskelligartet materiale trækkes individuelt og efter 90 minutters forberedelse fremlægger eksaminanden på 7–10 minutter to til tre problemstillinger og svarer på dem efterfulgt af dialog, derefter votering. Offentliggørelse af karakteren er ikke beskrevet. I alt 30 min pr eksamination	1. Læreplan 4.2
Aktører	1. Eksaminand, eksaminator og evt censor har dialog 2. Votering mellem eksaminator og censor 3. Et anonymt <i>man</i> udarbejder materialet. I praksis forventes det at være eksaminator	1. Læreplan 4.2 2. Vejl. s. 31 3. Vejl. s. 28–29

Uddybning af konstruktaspekterne *Kriterier for eksamen* og *Testmetode* på fagniveau – Historie A alment gymnasium (STX)

De ekspliciterede faglige mål udgør officielt de formuleringer, der retviser bedømmelsen:

Eleverne skal kunne:

1. redegøre for centrale udviklingslinjer og begivenheder i Danmarks, Europas og verdens historie
2. redegøre for sammenhænge mellem den lokale, nationale, regionale, europæiske og globale udvikling
3. analysere eksempler på samspillet mellem mennesker, natur, kultur og samfund gennem tiderne
4. skelne mellem forskellige typer af forklaringer på samfundsmæssige forandringer og diskutere periodiseringsprincipper
5. reflektere over samspillet mellem fortid, nutid og fremtid samt over mennesket som historieskabt og historieskabende
6. anvende en metodisk-kritisk tilgang til at udvælge og analysere historisk materiale, herunder eksempler på brug af historie
7. opnå indsigt i, hvordan historiefaget kan medvirke til at forstå og løse problemer i nutiden
8. formulere historiske problemstillinger og relatere disse til elevernes egen tid
9. formidle og remediere historiefaglige problemstillinger mundtligt og skriftligt og begrunde de formidlingsmæssige valg

10. behandle problemstillinger i samspil med andre fag
 11. demonstrere viden om fagets identitet og metoder.
- (Undervisningsministeriet, 2017, s. 1)

De mange læringsmål viser samme bredde som uddannelsens overordnede formål og indledes med verber, der antyder forskellige taksonomiske niveauer mere eller mindre loyalt udformet efter Blooms taksonomi (Bloom, 1956; Eskelund Knudsen, 2015). Skal det i en eksamen være muligt at udprøve alle disse læringsmål, skal eksamensdesignet være bredt, hvis ikke eksamen skal risikere at blive udfordret af Messicks begreb om construct underrepresentation. Kun visse af læringsmålene kan udprøves på baggrund af ét tema, 3–5 siders kildemateriale og et oplæg med efterfølgende dialog. De er formuleret generelt, det vil sige ikke i forhold til et konkret udleveret materiale. Der er risiko for, at et bestemt eksamenssæt kun kan dække en vis del af læringsmålene. Vejledningen ekspliciterer denne risiko ved at formulere, at den mundtlige testmetode medfører, at det ”normalt ikke [vil] være muligt for eleven/kursisten at demonstrere en udfoldet opfyldelse af alle faglige mål, men det *tilstræbes* at der gennem materialesæt og eksamination åbnes for målopfyldelse af et betydeligt antal heraf” (Undervisningsministeriet, 2018, s. 31, min kursivering)¹⁰. Vejledningens fortolkning af formuleringen: ”Materialesættene, der indgår som grundlag for prøven, skal tilsammen *i al væsentlighed* dække de faglige mål og de gennemførte forløb” er, at ”korterevarende faglige aktiviteter (temadage, optakter til studietur med andre fag, m.m.) ikke behøver at indgå ved eksamen, ligesom man i særlige tilfælde kan undlade at lave materialesæt til et forløb, der af forskellige årsager ikke egner sig hertil” (Undervisningsministeriet, 2018b, s. 30). Undervisningsforløb og ikke-faglige mål kan være fritaget i designet af det *samlede* antal materialesæt. I eksemplet på karakterbeskrivelser fremgår der diskursive henvisninger til læringsmål 1,2,3,5,6,9 og 11 i beskrivelsen af topkarakteren, hvorimod læringsmål 4,7,8 og 10 ikke fremstår tydeligt. Tillægsordene *præcis*, *velstruktureret* og *udtømmende* beskriver topkarakteren, og de står strengt taget ikke i læringsmålene og må bero på en fortolkning af *formidling* i læringsmål 9. Det er muligvis kvaliteter ved en mundtlig fremstilling, men de er ikke opført i de 11 læringsmål, og der er risiko for construct-irrelevant variance.

Analyse af Interpretation/Use Argument

Læsningen af de fire centrale dokumenter, loven, eksamensbekendtgørelsen og eksemplet historiefagets læreplan og vejledning viser, at slutningsformen generalisering står centralt både på uddannelses- og fagniveau, hvilket er almindeligt for tests, men omfanget er bemærkelsesværdigt. Generaliseringslutningerne står

¹⁰ Danmark har ingen uddannelseskrav til eksaminatorer i forbindelse med design af eksamensmateriale eller eksamination.

ikke ekspliciteret i styredokumenterne, men må være implicit til stede for, at det kan hævdes, at eksamen er relevant for det meget brede formål med uddannelsen. En fremanalyseret slutning må være, at eftersom uddannelsens formål ikke udprøves generisk i sig selv, må uddannelsens formål udprøves i fagene, fag der er samplet efter uigennemskuelige principper. Når der generaliseres fra fagniveau til overordnet uddannelsesformål, kan fagene udskiftes; det gør ikke nogen forskel, om eksaminanden er testet i det ene eller det andet fag med generaliseringshjemmelen: Hvad der gælder for det ene fag, gælder for alle fag – det gælder for uddannelsen. *Slutning 1: Uddannelsens formål udprøves lige godt i alle fag.* Dette er paradoksalt, eftersom formålet med eksamen i eksamensbekendtgørelsen sidestiller uddannelsesniveau og fagniveau, men i realiteten måles det enkelte fags læringsmål og *derigennem* uddannelsens formål. Det samme samplingsprincip ses inde i faget: Materialet/spørgsmålet samples, så den præsentation, der vurderes i den mundtlige eksamen, antages at være generaliserbar til præstationer, som eleven ville levere ved et andet materiale/spørgsmål. *Slutning 2: Eksamenssæt i faget er lige svære og dækker det, testen intenderer at måle.* På den måde vil mundtlig eksamen i historie med træk af et sæt om ”Ideologiernes kamp” kunne udgøre en betragtelig kerne i det samlede resultat af studentereksamen. Det samme gælder for eksaminandens øvrige mundtlige udtræksfag. Det testen intenderer at måle, beskrives i 11 læringsmål, som testmetoden normalt ikke vil kunne dække, jf. vejledningen, og det vil kunne læses som generalisering fra et læringsmål til et andet: *Slutning 3: Elevens præsentation vurderet ud fra nogle læringsmål kan generaliseres til de andre.*

Diskussion af slutningerne

Analyseresultaterne viser, at konstruktet står uklart både på uddannelses- og på fagniveau – i det omfang fundene for historie kan overføres til andre fag. Som konsekvens heraf er udgangspunktet for at validere eksamensresultaterne svagt. Det betyder, at spørgsmål om construct underrepresentation og construct-irrelevant variance ikke kan afgøres. Slutninger om det eksisterende systems sammenhæng er ikke ekspliciteret af testdesigneren og lever ikke op til Standards eller Kanes første krav om eksplicit transparens. Muligvis skyldes det, at testdesigneren ved mundtlige studentereksaminer ikke kan identificeres, da testdesign både foregår centraliseret i form af læringsmål udformet som heuristik og lokal udvælgelse i den praktiske fortolkning af denne heuristik. Systemets kohærens kan ikke vurderes pga. manglende transparens. Det tredje krav om rimelighed må vurderes efter, hvor acceptable man finder slutningerne om generalisering. De kan i princippet undersøges empirisk og diskuteres, og en forsigtig vurdering lyder, at der er tale om Kanes ”weak ”hidden” assumptions” (Kane, 1992, s. 528). Slutning 1 viser en spænding mellem uddannelsens formål, der ikke i sig selv udprøves, og faget, hvor eksamen foregår. Udtræk mellem

fagene bringer principielt det enkelte fag ud af funktion. Rationalet bag udtræksordningen er ikke ekspliciteret i styredokumenterne. Den norske Eksamensgruppa skriver, at der mangler information om dens begrundelser og overvejer pragmatiske, ressourcemæssige og fagpolitiske årsager (Utdanningsdirektoratet, 2019b, 7.1.1). I litteraturen henvises også til elevmotivation som baggrund:

Vissheten om at det kan komme en eksamen i faget skal bidra til å holde elevene på «tå hev». Eksamen har således også en rolle som et ytre press – som skal bidra til at elevene arbeider godt i alle fagene i skolen. (Tveit & Olsen, 2018, s. 23; se også Gipps, 2012, s. 12; Utdanningsdirektoratet, 2019b, 7.1.1)

Hvis antagelsen er, at elever (og lærere?) præsterer en større arbejdsindsats ved udtræk af eksaminer trods et sideløbende årskaractersystem til at motivere kontinuerligt arbejde (se Hovdhaugen et al., 2018), efterstår et empirisk arbejde i at kortlægge sådan praksis. Standards skriver, at tests, der begrundes i motivationsantagelser, må belægges:

Educational tests, for example, may be advocated on the grounds that their use will improve student motivation to learn or encourage changes in the classroom instructional practices and holding educators accountable for valued learning outcomes. Where such claims are central to the rationale advanced for testing, the direct examination of testing consequences necessarily assumes even greater importance. Those making the claims are responsible for the evaluation of the claims. (Standards, 2014, s. 20)

Sidst kunne udtrækssystemet i sin nuværende form tænkes at løfte formålet evaluering af undervisningen eller accountability (Gipps, 1994, s. 11), dvs. at sampling på landsplan i alle fag giver indtryk af gymnasiernes leverance til samfundet. Den enkelte elevs kompetencer flyttes da væk fra eksamens centrum. Disse formål er ikke ekspliciteret i eksamensbekendtgørelsen og strækker antagelser om generalisering af den enkelte elevpræstation endnu længere.

Slutning 2 om udtræk mellem eksamensmaterialer/spørgsmål kan være holdbar, men eftersom sværhedsgraden af eksaminationsmaterialet udformes lokalt og ikke undersøges yderligere end censors godkendelse, er det sårbart for praktiske omstændigheder i forbindelse med tid allokeret til udarbejdelsen. Den lokale udformning af eksamensmaterialet giver sandsynlighed for en nær sammenhæng mellem undervisning og eksamination, og censorsystemet giver mulighed for kollegial indsigt i andre eksaminators praksis. Møde mellem bedømmere kan være både en styrke og en svaghed, eftersom fagpersoner får anledning til at se og diskutere håndgribelige eksempler på, hvordan elevpræsentationer har kvalitet – afhængigt af faktorer som hvor velfungerende interaktionen mellem bedømmerne er. Dette klargør ikke nødvendigvis konstruktet. Slutning 3 om læringsmålenes tilfældighed er ligeledes interessant her, da bedømmerne må have en vis enighed om tegn på kvalitet eller diskussion om rimeligheden af fortolkningen af et omfattende antal læringsmål, der ikke er realistiske at udprøve i praksis.

Diskussion af generaliseringslutningers berettigelse

Generaliseringen af testresultatet bygger på antagelsen om, at eksaminanden ville levere en tilsvarende præstation, hvis man testede i et andet, tilsvarende materiale, altså at der er substantiel lighed i de mulige samplinger fra "the universe of observations" (Kane, 1992, s. 529). En eksaminands præstation under overskriften "Ideologernes kamp" kan generaliseres til de præstationer, eleven ville give under alle de andre overskrifter i historiefaget. En mulighed er, at eksaminator og censor i den mundtlige eksamen forholder sig til netop den præsentation og dialog om "Ideologernes kamp", de har med eksaminanden og ikke vurderer, hvorvidt præstationen repræsenterer andre dele af faget endsige andre aspekter af uddannelsens formål, som nødvendigvis er indeholdt i udtrækssystemet. I udtrækssystemet opstår der spænding mellem de materiale, faglige indholdsdimensioner og formale, generelle evner (Klafki, 2002, s. 15; Klafki, 1983), da fagets substantielle indholdsdimensioner ikke kan sidestilles i det samlede eksamensresultat.

At antage at den konkrete præstation kan generaliseres til andre dele af faget, er mere nærliggende, end at en præstation i historie kan repræsentere en præstation i fx matematik. Klafki formulerer hverdagserfaring, der indeholder generaliserings- og konsekvensproblemer der kan opstå for eleverne i et udtrækseksamenssystem, der generaliserer fra det ene fag til andre:

Der findes i pædagogiske sammenhænge en almindelig erfaring, der på slående vis tydeliggør den funktionelle dannelsesteori og dens forudsætningers pædagogiske utilstrækkelighed. Vi gør dag efter dag den iagttagelse, at f.eks. en elev, der i matematik udviser evner til at tænke i relationer, måske på ingen måde besidder denne færdighed inden for den sproglige sfære, at en anden, der udmærker sig med en stor fantasi hvad angår kunstnerisk skaben, virker udtalt fantasiløs, stillet over for den opgave at forsøge at udkaste hypoteser, der kan forklare simple fysiske sagsforhold [...] hvad "fantasi", "relations-tænkning", "iagttagelsesevne" osv. er, det er åbenbart afhængigt af strukturen af indholdet, der tænkes, udkastes som fantasiforestilling, iagttages som "genstande". (Klafki, 1983, s. 49)

De formale kompetencer kan altså ikke nødvendigvis overføres på forskelligartet materiale på tværs af fag. Der er både materiale og formale elementer i historiefagets 11 læringsmål, om end de generelle formale kompetencer som kommunikative, interaktionelle kompetencer ikke fylder nær så meget som materiale historiefaglige. Et stærkere argument for, at uddannelsens formål udprøves gennem historiefaget, kunne muligvis bygges ved at anerkende sådanne formale kompetencer i højere grad.

Diskussion af eksamen i didaktik- og curriculumtraditioner

Tilbage står spørgsmålet om, hvorvidt det er kommensurabelt at læse danske mundtlige eksaminer forankret i didaktiktradition med amerikansk assessment-teori. Den didaktiske tradition retter fokus mod læreplanen (Gundem, 1997; Imsen, 2006; Klafki, 1983; Westbury et al., 2000). Smagorinsky beskriver didaktisk og curriculum-traditionelt sammenstød i et konkret *hybrid classroom*, ”that does not reach the goal of either school of teaching” (Smagorinski et al., 2002, s. 203), hvilket kunne inspirere til at anse de danske mundtlige eksaminer for *hybrid examinations*, fordi et centraliseret system med upræcist konstrukt er åbent for eksaminators og censors fortolkning af kvalitet. Struktur-aktørforholdet kan være mere eller mindre dynamisk (Imsen, 2006, s. 248). Et spørgsmål fra en assessmentteoretisk vinkel til det didaktiske fokus på læreplanen kunne være: Kan der sættes lighedstegn mellem undervisningens ideelle læringsmål og eksamenskonstruktionen? Er det identisk at opstille mål for undervisningen og at genkende det pågældende mål ved eksaminandens mundtlige eksamenspræstation? Hvordan stiller det eleven, for hvem eksamen har konsekvenser? Didaktiktraditionen fordrer den reflektive lærer at definere mål og midler i undervisnings-sammenhæng, og muligvis er eksamens testfunktion irrelevant for lærergerningen, et administrativt fænomen der ligger uden for den egentlige skolevirkingsområde? Eksamensbekendtgørelsen formulerer, at det, der kvalificerer til at fungere som censor, er, at man er kompetent til at varetage et undervisningsforløb. Den didaktisk velreflekterede lærer definerer mål for eksamen qua sin måldefinition for undervisningen. Forestillingen om målet og målets konkrete udtryk smelter sammen.

Komplikationerne ved at identificere og genkende målestokken har assessmentlitteraturen reflekteret over siden 1950'erne. Gundem kombinerer didaktik- og curriculumtraditionelle tilgange (Gundem, 1997). Klafki og assessmentteoretikerne er enige på en række vigtige punkter: Eksamen handler grundlæggende om emancipationsmuligheder for den enkelte elev som alternativ til reproducerede sociale arvelige strukturer (Klafki, 2002, s. 242). Både Klafki og assessmentteoriene konstaterer, at eksamen er en administrativ kendsgerning og har ikke bedre bud på en passende selektionsmekanisme (Klafki, 2002, s. 267). Begge har fokus på konsekvensperspektivet i en social kontekst (Klafki, 2002, s. 271). Det må tydeliggøres, hvad der måles, det må begrundes, og det inkluderer omfangsrige refleksioner og diskussioner mellem praktikere, forskere og elever (Klafki, 2002, s. 263–264). Disse refleksioner kunne inspireres af assessment-teori. Klafki foreslår at flytte fokus fra ”objektiverede eller direkte konstaterbare resultater af elevens anstrengelser” til ”åndelige processer, fx gennemførelsen af kommunikation i undervisningen, udviklingen af en kritik”, selvom det er ”langt sværere at udvikle objektiverbare kriterier herfor” (Klafki, 2002, s. 260). Ingen vil hævde, at konstruktet kan beskrives udtømmende en gang for alle. Det er accepteret i assessmentteori, at ”the validation process never ends” (Standards,

2014 s. 21), fordi der altid vil være nye formuleringer af konstrukter og slutninger til validering. Assessmentlitteraturen, der er produceret i curriculum-traditionelle angelsaksiske lande, har med Standards for Educational and Psychological Testing og argumentationsbaserede tilgange som Kanes ekspliciteret, at fortolkning bør sættes i centrum for validitetsstudier af tests. Hvis kvalitet og konstrukt materialiserer sig for en lærer i didaktisk tradition ved genkendelse, når den mødes, findes konstruktet tavst, og er ikke operationaliserbart for elev/forældre/samfund. Eksaminanden må fortolke det implicite konstrukt for at undgå dårlige eksamensoplevelser og optimere egne fremtidige muligheder for et frit uddannelsesvalg.

Konklusion

Danske mundtlige eksaminers konstrukt står ikke veldefineret, når eksamenssystemets centrale styredokumenter læses med assessmentteoretiske begreber. Slutninger om omfattende generalisering af eksamensresultater må fortolkes frem på tværs af dokumenterne. Med et usikkert referencepunkt er validering af elevernes eksamensbesvarelser svært. Styredokumenternes centraliserede funktion er, at elever skal ligestilles: "The reason for uniform procedures is to allow comparability between the results of all students, who may take the tests in different places" (Harlen, 2007, s. 17). Den ensrettede procedure i det danske system er, at alle elever er underlagt samme ikke-ensrettede procedure. En norsk student kalder det "en bingo-eksamen" (Kvifte, 2011, s. 149): Man kan være heldig eller uheldig med sine eksamensfag, med sit eksamenssæt og med sine bedømmere.

Selvom styredokumenterne ikke er tilstrækkelige fra et assessmentteoretisk synspunkt, kan lokal eksaminationspraksis på skolerne godt have ekspliciterede rationaler og konstruktforståelser. Ved at lade konstruktdefinition være en kontinuerlig ekspliciteringsproces blandt fagfolk, kan assessmentteori anvendes som inspiration for eksaminationsspørgsmål i en dansk didaktisk tradition. Praksis kan i fremtidig forskning undersøges med spørgsmål som: Hvad tænker eleverne, at de testes i? Hvordan genkender en eksaminator kvalitet ved mundtlig eksamen? I hvilken grad er der transparens og enighed om kvalitet ved en eksaminands mundtlige eksamenspræstation?

Om forfatteren

Julie Marie Isager er ph.d.-studerende i Uddannelsesvidenskab ved Syddansk Universitet. Julie forsker i mundtlige eksaminer og elevers forståelse og forberedelse til eksamen i og uden for undervisningen.

Institutionel tilknytning: Institut for kulturvidenskaber, Syddansk Universitet, Campusvej 55, 5230 Odense M, Danmark.

E-post: juliei@sdu.dk , julieisager@gmail.com

Referencer

- Betænkning over Forslag til lov om de gymnasiale uddannelser (2016). Nr. 2016/1 BTL 58. <https://www.retsinformation.dk/Forms/R0710.aspx?id=185774>
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain* (17. pr.). New York.
- Borger, L. (2018). *Investigating and validating spoken interactional competence: Rater perspectives on a Swedish national test of English*. Acta universitatis Gothoburgensis.
- Bøhn, H. (2016). *What is to be assessed? Teachers' understanding of constructs in an oral English examination in Norway*. Doktorafhandling, Universitetet i Oslo. <https://www.duo.uio.no/handle/10852/53229>
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Crooks, T. J., Kane, M. T. & Cohen, A. S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265–286. <https://doi.org/10.1080/0969594960030302>
- Danmarks Evalueringsinstitut (2003). *Eksamensformer i det almene gymnasium*. Danmarks Evalueringsinstitut.
- Dobson, S. (2008). Theorising the academic viva in higher education: The argument for a qualitative approach. *Assessment & Evaluation in Higher Education*, 33(3), 277–288. <https://doi.org/10.1080/02602930701293272>
- Eksamenbekendtgørelsen (2016). *Bekendtgørelse om prøver og eksamen i de almene og studieforberedende ungdoms- og voksenuddannelser*. Undervisningsministeriet. BEK nr 343 af 08/04/2016. <https://www.retsinformation.dk/forms/R0710.aspx?id=179722>
- Erickson, G. & Gustafsson, J.-E. (2014). Bedömningens dubbla funktion—För lärande och likvärdighet. I U. P. Lundgren, R. Säljö & C. Liberg (red.), *Lärande, skola, bildning: Grundbok för lärare* (3. rev. och uppdaterade utg., s. 559–589). Natur & kultur.
- Eskelund Knudsen, H. (2015). Historiefagets forhold til redegørelse. *Nordidactica: Journal of Humanities and Social Science Education*, 2015(1), 47–63.
- Gipps, C. V. (2012). *Beyond testing: Towards a theory of educational assessment*. Falmer Press.
- Gundem, B. B. (1997). Läroplansarbete som didaktisk verksamhet. I M. Uljens (red.), *Didaktik: Teori, reflektion och praktik* (s. 246–267). Studentlitteratur.
- Harlen, W. (2007). *Assessment of learning*. SAGE Publications.
- Haue, H., Nørr, E. & Skovgaard-Petersen, V. (1998). *Kvalitetens vogter: Statens tilsyn med gymnasieskolerne 1848–1998*. Undervisningsministeriet.

- Hovdhaugen, E., Prøitz, T. S. & Seland, I. (2018). Eksamens- og standpunkt karakterer – to sider av samme sak? *Acta Didactica Norge*, 12(4), Art. 17.
<https://doi.org/10.5617/adno.6276>
- Huxham, M., Campbell, F. & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment & Evaluation in Higher Education*, 37(1), 125–136. <http://dx.doi.org/10.1080/02602938.2010.515012>
- Imsen, G. (2006). Allmenn didaktikk og fagdidaktikk—Mellom dannelse og utdanningspolitikk. I S. Ongstad (red.), *Fag og didaktikk i lærerutdanning: Kunnskap i grenseland* (s. 243–257). Universitetsforlaget.
- Joughin, G. R. (2003). *Oral assessment from the learner's perspective: The experience of oral assessment in post-compulsory education*. Ph.D., Griffith University, Australia.
<http://search.proquest.com/docview/305268419/abstract/9FE883360C4A4F09PQ/1>
- Kalthoff, H. (2013). Practices of Grading: An Ethnographic Study of Educational Assessment. *Ethnography and Education*, 8(1), 89–104.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2006). Validity. I R. L. Brennan, National Council on Measurement in Education & American Council on Education, *Educational measurement* (4. udg., s. 17–64). Praeger Publishers.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Klafki, W. (2002). *Dannelsesteori og didaktik: Nye studier*. Klim.
- Klafki, W. (1983). *Kategorial dannelse og kritisk-konstruktiv pædagogik*. Nordisk Forlag.
- Kvifte, B. H. (2011). *Muntlig eksamen sett fra studentperspektiv: En undersøkelse blant lærerstudenter ved Høgskolen i Østfold*. Rapport 2011:3, Høgskolen i Østfold.
<https://hiof.brage.unit.no/hiof-xmlui/bitstream/handle/11250/148642/Hefte3-2011.pdf?sequence=1&isAllowed=y>
- Lov om de gymnasiale utdannelse (2016). *LOV nr 1716 af 27/12/2016*.
<https://www.retsinformation.dk/eli/lta/2016/1716#id1e65fa4c-cb8b-48b4-905c-9610769e2c34>
- Maugesten, M. (2011). Muntlig eksamen. En analyse av åtte studenters forståelse av muntlig eksamen i matematikk. *Norsk pedagogisk tidsskrift*, 95(04), 260–272.
- Messick, S. (1989). Validity. I R. L. Linn (red.), *Educational Measurement* (s. 13–103). Macmillan.
- Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning. *American Psychologist*, 50(9), 741–749. <https://ovidsp-dc1-ovid-com.proxy1-bib.sdu.dk/sp-3.33.0b/ovidweb.cgi?QS2=434f4e1a73d37e8c6114a63cc85fea098b58a4e129984b8d8ba371f5a5b964f938de9fa14dfdb51f94eade4d2d34bfceb4d535fad7f54bfa0fbbf2921f5174d8146e7d5adbe1baa67408e26e239809bf25cd789c05bbf0f8cc9edfad456e500925c81922341b8e984aa63accf82249f3d87129f88f2111800e78dc266050673be48262925124a5f491805ef06e6af8001fe900e8d263c047fc63c69e553c38c377bd59fd54fc9341b11ebec8b97ee56d9612a763c6564194c3c3bbf93a7d964feb6bfc9571b27c27606cfc9d114ee9803cf79e206b63a787153b3043861750f72a361011e9c4e8c0807adbdc4095e97001f0ff816ae727116d5648d440a581603a845b57e65f9fd35e5ef175333838ca159217ce000499b67890f666b77ef977e83c576ec2dd6f5179dfab1cc5eb93>
- O'Neill, L. D., Lykkegaard, E. & Kulasageram, K. (2019). Intended and unintended test constructs in a Multiple-Mini admission Interview. A validity study. *Dansk Universitetspædagogisk Tidsskrift*, 14(26), 66–81.

- Palmér, A. (2010). *Att bedöma det muntliga: Utvärdering av ett delprov i gymnasieskolans nationella kursprov, Svenska B*. Rapport, Uppsala universitet.
<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-144126>
- Pearce, G. & Lee, G. (2009). Viva Voce (Oral Examination) as an Assessment Method: Insights From Marketing Students. *Journal of Marketing Education*, 31(2), 120–130.
<https://doi.org/10.1177/0273475309334050>
- Perelman, C. & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. (Reprinted). University of Notre Dame Press.
- Rolin, A. (2013). Argumentasjon og interaksjon i muntlig eksamen. *Acta Didactica Norge*, 7(1), Art. 16. <https://doi.org/10.5617/adno.1120>
- Skov, S. (2013). Nye læringsmål kræver nye eksamensformer. *Dansk Universitetspædagogisk Tidsskrift*, 8(14), 66–75.
- Smagorinski, P., Lakly, A. & Johnson, T. S. (2002). Acquiescence, Accomodation, and Resistance in Learning to Teach within a Prescribed Curriculum. *English Education*, 34(3), 187–213.
- Standards (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association, American Psychological Association, National Council on Measurement in Education & Joint Committee on Standards for Educational and Psychological Testing.
- Tofteskov, J. & Leth Andersen, H. (2016). *Eksamen og eksamensformer: Betydning og bedømmelse* (2. udg.). Samfundslitteratur.
- Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press.
- Tveit, S. & Olsen, R. V. (2018). Eksamens mange roller i sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring. *Acta Didactica Norge*, 12(4), Art. 18.
<https://doi.org/10.5617/adno.6381>
- Undervisningsministeriet (1964). Bekendtgørelse om eksamensordningen og karaktergivning m.v. ved studentereksamen og i gymnasiet.
https://library.au.dk/uploads/tx_lfskolelov/1964-04-20_123.pdf
- Undervisningsministeriet (2008). *Prøve- og eksamenssystemet—Udfordringer og muligheder* Rapport fra en arbejdsgruppe nedsat af undervisningsministeren.
<http://static.uvm.dk/publikationer/2008/eksamenssystemet/helepubl.pdf>
- Undervisningsministeriet (2017). *Historie A/B, STX/hf-enkeltfag læreplan*.
<https://www.uvm.dk/gymnasiale-uddannelser/fag-og-laereplaner/laereplaner-2017/stx-laereplaner-2017>
- Undervisningsministeriet (2018). *Historie A/B, STX/hf-enkeltfag vejledning*.
- Utdanningsdirektoratet (2019a). *Kunnskapsgrunnlag for evaluering av eksamensordningen*. Utdanningsdirektoratet. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/Kunnskapsgrunnlag-for-evaluering-av-eksamensordningen/>
- Utdanningsdirektoratet. (2019b). *Vurderinger og foreløpige anbefalinger fra eksamensgruppa*. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/vurderinger-og-forelopige-anbefalinger-fra-eksamensgruppa/>
- Westbury, I., Hopmann, S. & Riquarts, K. (2000). *Teaching as a reflective practice: The German didaktik tradition*. Lawrence Erlbaum.

Bilag: Litteratursøgning

Litteratursøgning med formålet at identificere dansk forskning i studentereksamensvaliditet på dansk og engelsk. Søgestrengen er lavet inspireret af PICO og skal kunne favne assessment- og validitetsbegreber og er søgt holdt åben ved at udelade reliabilitet. Irrelevante kilder (IR) vurderes efter kriterierne: Uddannelse, gymnasieniveau, testen er summativ afgangseksamen, der tester læring, der skal indgå et mundtligt element, og være en generel test, dvs. at der ikke fokuseres på grupper af elever med bestemte behov.

Dansk søgestreng: (Danmark OR dansk) AND (validitet OR valid OR gyldighed) AND (eksamen OR test OR "mundtlig eksamen" OR assessment OR studentereksamen) AND (gymnasieelev OR eksaminand OR gymnasiasst OR student)

Engelsk søgestreng: (Denmark OR danish OR danes) AND (validity OR valid OR accuracy OR validation) AND (exam OR test OR assessment OR "summative assessment" OR "oral exam" OR "oral test" OR examination OR "final exam") AND ("high school student" OR "secondary school student" OR "high schoolers" OR "secondary-school student" OR "secondary school pupils")

Database	Søgeresultat dansk	Søgeresultat engelsk
Scopus	0	3 IR
Web of Science	0	1 IR
Idunn	16 IR, 1 om censur	2 IR
Academic Search Premiere + Eric	13 IR	1 IR
Encyclopedia of Language and Education	0	0
Google Scholar	12500	4270
SDU Summon	57079	9697
Forskningsdatabasen.dk	6 IR	4 IR