Linköping Studies in Behavioural Science No. 256

Varierade bedömningspraktiker i samhällskunskap

Validitet och agens i lärares beslut om bedömningars syften, innehåll och metoder

Tobias Jansson

LINKÖPINGS UNIVERSITET

Linkoping studies in Behavioural Science No. 256

Varierade bedömningspraktiker i samhällskunskap

Validitet och agens i lärares beslut om bedömningars syften, innehåll och metoder

Tobias Jansson

Institutionen för beteendevetenskap och lärande,
Avdelningen för pedagogik och didaktik
Utbildningsvetenskap
Linköpings universitet, 581 83 Linköping
Linköping 2024

Detta verk är licensierat under en Creative Commons Erkännande 4.0 Internationell Licens.
https://creativecommons.org/licenses/by/4.0/deed.sv

© Tobias Jansson, 2024

Tryck: LiU-Tryck, Linköping 2024

ISBN 978-91-8075-446-0 (print)
ISBN 978-91-8075-447-7 (PDF)
https://doi.org/10.3384/9789180754477
ISSN 1654-2029

Omslagsbild: Martin Barraud, iStock

Abstract (English)

This thesis explores Civics teachers' assessment practices regarding aims, content, skills and methods, and factors that may affect teachers' assessment practices. This was examined through thematic analyses of interviews with a total of 19 upper secondary school teachers and of their assignments. The studies showed a variation, both between the teachers, and that individual Civics teachers varied their assessments. The teachers in the study had varied purposes, but a performative discourse, emphasising the grading criteria, resulted in a summative practice focusing on evidence for grading. On the other hand, the teachers expressed a high degree of agency regarding content, skills and methods. However, a subject tradition affected what content is tested. Historically, teachers have mainly tested factual knowledge in written tests, but the teachers in these studies varied their methods and emphasized reasoning skills, which is in line with the open character of the subject.

The focus on grading criteria is positive for the validity of grading, but validity in relation to formative purposes can be problematized. However, the emphasis on summative assessment entails a risk of an instrumental view of school, focusing on grades rather than learning. Variation between teachers also risks problems arising regarding equal assessment, where students can learn different things and develop different skills. On the other hand, variation may help more students to succeed, as different methods suit different students.

The teachers were socialised into assessment practices, which have been shaped primarily by their experiences, rather than teacher education. However, education in assessment could increase teachers' assessment competence. Assessment is difficult and complex, and teachers need to balance different purposes, choose appropriate methods in relation to those purposes and what content and skills are to be tested, as well as deal with the various consequences of assessments. It is important work, in which Civics teachers may achieve a high degree of agency and professionalism.

Keywords: Classroom assessment, assessment practice, teacher agency, performativity, validity, Civics, social studies, upper secondary school teachers, thematic analysis

Abstract (svensk)

I denna avhandling undersöks samhällskunskapslärares bedömningspraktiker vad gäller syften, innehåll, förmågor och metoder, samt faktorer som kan påverka lärares bedömningspraktiker. Detta undersöktes genom tematiska analyser av intervjuer med totalt nitton gymnasielärare och deras bedömningsuppgifter. Studierna visade en variation, både mellan lärarna och att enskilda samhällskunskapslärare varierade sina bedömningar. Lärarna i studierna hade varierade syften, men en rådande performativ diskurs, med betoning på betygskriterierna, innebar en summativ praktik med fokus på att ha underlag för betyg. Däremot uttryckte lärarna en hög grad av agens gällande innehåll och förmågor som testades och med vilka metoder. Samtidigt påverkade en ämnestradition vilket innehåll som testades. Historiskt har lärare främst testat faktakunskaper i skriftliga prov, men lärarna i studierna varierade metoderna och betonade resonerande förmågor, vilket är i linje med ämnets öppna karaktär.

Fokuseringen på betygskriterierna är positiv för validiteten gällande betygsättning. Däremot kan validiteten i relation till formativa syften problematiseras. Betoningen på summativ bedömning innebär dock en risk för en instrumentell syn på skolan med fokus på betyg snarare än lärande. Variationen mellan lärare riskerar också att leda till problem med likvärdig bedömning där elever lär sig olika saker och utvecklar olika förmågor. Samtidigt hjälper variation fler elever att lyckas då olika metoder passar olika elever.

Lärarna har socialiserats in i en bedömningspraktik, vilken främst har formats av deras erfarenheter, snarare än lärarutbildningen. Utbildning i bedömning skulle dock kunna öka lärares bedömarkompetens. Bedömning är svårt och komplext, där lärare behöver balansera olika syften, välja ändamålsenliga metoder i relation till syften och vilket innehåll och vilka förmågor som ska testas, samt hantera bedömningars olika konsekvenser. Det är ett viktigt arbete, i vilket samhällskunskapslärarna har möjlighet till stor agens och professionalitet.

Nyckelord: bedömning, bedömningspraktik, läraragens, performativitet, validitet, samhällskunskap, gymnasielärare, tematisk analys

Förord

Det har varit en lång resa fram till nu när denna avhandling äntligen är färdigskriven. På ett sätt skulle man kunna säga att resan började redan under min lärarutbildning. Mitt intresse för konstruktionen av bedömningsuppgifter väcktes av min metodiklektor i historia, Tommie Lundquist, utifrån hans konstruktiva kritik av mina första tragglande försök att konstruera ändamålsenliga prov i relation till det innehåll och de förmågor som skulle testas. Jag minns fortfarande när jag hösten 1995 i ett seminarium stolt visade upp ett prov som jag hade konstruerat under min praktik. Provet handlade om imperialismen och innehöll flera frågor om bland annat orsaker och konsekvenser och skulle inte bara testa faktafrågor. Trodde jag. Tommie tittade lurigt på mig, kliade sig på hakan och frågade: ”Men stod inte orsakerna i boken och hade du inte gått igenom konsekvenserna på tavlan?” Jo, så var det ju. ”Men då testar ju ditt prov bara vad eleverna kan komma ihåg!” Detta blev en väckarklocka för mig och hjälpte mig att som lärare tänka till lite extra när jag utformade olika bedömningsuppgifter. Ett stort tack till Tommie för detta!

Det dröjde dock till hösten 2008 innan jag fick fördjupa mig om bedömning, när jag påbörjade mina forskarstudier inom ramen för *Forskarskolan för lärare i historia och samhällskunskap* vid Karlstads universitet. Det var då som min första delstudie, licentiatuppsatsen, genomfördes. Valet av ämne var för mig redan från början självklart. En av forskarskolans inriktningar var bedömning och det var detta jag ville fördjupa mig inom. Vid den här tidpunkten var det dock ämnet samhällskunskap som intresserade mig mer, eftersom det har en så resonerande karaktär.

Jag är tacksam för denna forskarutbildning, vilken lade grunden för min fortsatta karriär. Den möjliggjorde att jag fick anställning vid Linköpings universitet, där jag bland annat har fått undervisa om bedömning. Så småningom öppnades möjligheten att bygga på min licentiatuppsats inom ramen för min anställning. Jag vill rikta ett särskilt tack till mina avdelningschefer som givit mig denna möjlighet, först till Lotta Holme som erbjöd denna möjlighet och sedan till Margareta Engvall, Henrik Lindqvist och Åsa Järnvall som fortsatt ge mig möjlighet att forska.

Allra främst vill jag tacka mina handledare, Håkan Löfgren och Joakim Samuelsson, som med sina kloka kommentarer och frågor hjälpt mig framåt i processen och fördjupat min förståelse för vetenskap och vetenskapligt skrivande. Det har varit frustrerande ibland, men väldigt nyttigt. Jag är också tacksam för kommentarer från min diskutant vid slutseminariet, Christina Wikström, liksom till läsgruppen med Per Andersson och Robert Thornberg. Vidare tackar jag mina kollegor i nätverket *Forsk-*

ning om bedömning för läsning och kommentarer av texter. Jag riktar också ett tack till mina kollegor för uppmuntran, stöd och goda råd. Ett särskilt tack till Cecilia Björkhammer, Elisabeth Tenglet, Maria Sparf och Tove Mattsson för hjälp med referenslistan. Ett extra tack till min korridorgranne Elisabeth Eriksson, som ofta har varit mitt bollplank i olika frågor.

Jag tackar också alla de anonyma lärare som bidragit till denna studie genom att skicka material och ställa upp på intervjuer. Utan er hade det ju inte gått att genomföra denna studie. Jag hoppas att jag i denna studie har gjort er rättvisa, även om jag i min forskarroll också har behövt förhålla mig kritisk. Tack också till min hemgrupp i Brokyrkan för ert stöd och förbön.

Allra sist vill jag rikta ett extra stort tack till min älskade fru Paula. Tack för att du finns och för att du har stått ut med den frustration jag stundtals känt, men också för att du tagit sig tid och korrekturläst min avhandling. Det är dock mina barn, Ester och Hanna, som jag tillägnar denna skrift. Min förhoppning är att min avhandling ska bli ett litet bidrag till att utveckla de bedömningar som ni kommer att möta i skolan.

Rimforsa i januari 2024

Tobias Jansson

Avhandlingens delstudier

I avhandlingen ingår tre delstudier: en licentiatuppsats och två artiklar:

I. Jansson, T. (2011). *Vad kommer på provet? Gymnasielärares provpraxis i samhällskunskap* [Licentiatuppsats, Karlstad Universitet]. Karlstad University Studies: Studier i de samhällsvetenskapliga ämnenas didaktik, 5. https://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A380888

II. Jansson, T. & Löfgren, H. (2022). Balancing diverse purposes in Civics teachers' assessment decision making. *Scandinavian Journal of Educational Research, 1–15*. https://doi.org/10.1080/00313831.2022.2148280
Publicerad under CC-BY-NC-ND-licens

III. Jansson, T. (2023). Civics teachers' assessment practices in Swedish upper secondary schools. A qualitative study. *Journal of Social Science Education, 22*(2). https://doi.org/10.11576/jsse-5938
Publicerad under CC-BY-SA-licens

Delstudierna är publicerade med godkännande från förlag och tidskrifter.

Innehåll

1 INLEDNING 11
Syfte, frågeställningar och avgränsningar 14
2 BEDÖMNINGAR FÖRR OCH NU 17
Bedömningspraktiker i historien 17
Förändrad syn på bedömning 19
Det svenska betygssystemet 20
Ämnet samhällskunskap på gymnasiet 23
3 TEORETISK BAKGRUND 27
Blooms reviderade taxonomi 27
Validitet 28
Ramfaktorteori 31
Interna och externa faktorer 32
Läraragens 33
4 TIDIGARE FORSKNING 37
Olika typer av bedömningar 37
Lärares bedömningspraktiker 40
Bedömning i olika ämnen 42
Bedömning i samhällskunskap 45
Bedömningspraktikers kontextuella villkor 47
Bedömningars konsekvenser 51
Sammanfattning av tidigare forskning 54
5 METOD 55
Datainsamling 55
Den första datainsamlingen 55
Den andra datainsamlingen 56
Analyser av data 57
Forskningsetiska överväganden 58
Metoddiskussion 58
6 SAMMANFATTNING AV DELSTUDIERNA 61
Delstudie I (licentiatuppsats) – Vad kommer på provet? 61
Delstudie II (artikel) – Balancing diverse purposes in Civics teachers' assessment decision making 64
Delstudie III (artikel) – Civics teachers' assessment practices in Swedish upper secondary schools. A qualitative study. 67
7 DISKUSSION 73
Varierade syften 73
Varierade bedömningspraktiker 75
Ämnets karaktär möjliggör variation 77
Variation och validitet 79
Konsekvenser av varierade bedömningspraktiker 81

Samhällskunskapslärarnas agens 82
Studiens begränsningar och fortsatt forskning 85
Implikationer 87
8 SUMMARY 89
Background 89
Civics in Sweden *90*
Theoretical background 90
Bloom's Revised Taxonomy: *91*
Validity *91*
Theories about factors influencing teachers' work *92*
Earlier research 92
Methodology 93
Summary of the sub-studies 94
Study I – What is tested? (Licentiate Thesis) *94*
Study II – Balancing diverse purposes in Civics teachers' assessment decision making (Article) *95*
Study III – Civics teachers' assessment practices in Swedish upper secondary schools. A qualitative study (Article) *96*
Discussion 97
Limitations and implications *98*
REFERENSER 101
BILAGOR 117
Bilaga 1 – Kursplan Samhällskunskap 1a1 (2019) 117
Bilaga 2 – Kursplan i samhällskunskap 1b (2019) 119
Bilaga 3 – Missivbrev för den första datainsamlingen 121
Bilaga 4 – Missivbrev för den andra datainsamlingen 123

1 Inledning

Denna avhandling handlar om gymnasielärares bedömningspraktik i samhällskunskap. Det är viktigt att studera lärares bedömningspraktik då en stor del av lärares arbete handlar om att bedöma elevers utveckling (McMillan, 2018; Moss, 2013). Bedömning beskrivs som en av lärares svåraste, mest komplexa och viktigaste uppgifter (Moss, 2013). En definition av bedömning är att den handlar om att samla in, tolka och använda information för att stödja lärarens beslut för att dokumentera och främja elevers lärande (McMillan, 2018). I denna bedömningsprocess fattar lärarna flera olika beslut, till exempel om bedömningens syfte, vilket innehåll och vilka förmågor som ska bedömas, samt vilka metoder som ska användas för att mäta elevernas kunskaper. Alla dessa val och beslut formar en lärares bedömningspraktik.

Bedömningar kan ha olika syften (McMillan, 2018), till exempel att diagnostisera elevernas nuvarande kunskapsnivå, att möjliggöra återkoppling eller ge underlag för betyg. Beroende på lärares syften med bedömningen kan lärares val påverkas avseende vad de vill informera sig om relaterat till elevernas kunnande samt hur de ska samla in denna information. Lärare i svensk skola bedömer elevers prestationer i relation till de betygskriterier[1] som är formulerade i de olika kursplanerna. Betygskriterierna uttrycker vilka förmågor eleven ska visa i relation till ett specifikt innehåll.

Således behöver lärare besluta om vilket kunskapsinnehåll och vilka kognitiva förmågor som ska bedömas, samt med vilka metoder. Detta gäller oavsett om bedömningen är av mer summativ karaktär, det vill säga främst används för att summera vad eleven kan vid ett visst tillfälle (ofta som underlag för betyg) eller formativ, då den används för att främja lärande genom att läraren anpassar sin undervisning utifrån vad bedömningen visar eller ger eleverna återkoppling om styrkor och svagheter i prestationen. Det finns en hel del forskning om formativ bedömning och en del om summativ bedömning, främst om betygsättning (Forsberg & Lindberg, 2010; Lundahl m.fl., 2015; McMillan, 2013). Denna forskning har dock främst annat fokus än lärares bedömningspraktik i form av vad som bedöms och med vilka metoder. Om detta finns det relativt lite forskning, både internationellt och i Sverige (McMillan, 2013; Torrez & Claunch-Lebsack, 2014). Denna avhandling bidrar med kunskap om lärares bedömningspraktiker, vilket kan öka förståelsen för komplexiteten i lärares bedömningsarbete samt bidra till en utveckling av lärares bedömningspraktik.

[1] I grundskolans lägre årskurser benämns det ”kriterier för bedömning av godtagbara kunskaper”.

Hur lärare informerar sig om elevernas kunskaper, det vill säga vilka metoder de använder, är en fråga för den professionella läraren. Vilka möjligheter lärare har i praktiken, deras agens, när gör sina val, påverkas av olika kontextuella faktorer. Priestley m.fl. (2015) menar att läraragens påverkas av såväl lärares kapacitet och uppfattningar, som de kontextuella villkor som är för handen när läraren fattar sina beslut. Villkor som har visat sig påverka lärares bedömningspraktik är till exempel styrdokument (Wahlström & Sundberg, 2015), externa prov (exempelvis nationella prov) samt elever och föräldrar (McMillan, 2003; McMillan & Workman, 1999). Då bedömning är en så central del i lärares arbete är det viktigt med kompetens om hur bedömningar kan genomföras med hög kvalitet (McMillan, 2018). Forskning pekar dock på brister i lärares bedömarkompetens (Amani m.fl., 2021; Forsberg & Wermke, 2012) och att lärare kan ha svårt att förklara sin bedömningspraktik (McMillan & Nash, 2000). Studier av lärares bedömningspraktik kan bidra med kunskap som kan hjälpa lärare att resonera om sin bedömningspraktik.

Ämnets karaktär tycks också påverka lärares bedömningar (Biberman-Shalev m.fl., 2011; McMillan, 2001; Prøitz, 2013). Kunskaperna om ämnets påverkan på bedömning är dock begränsad, inte minst i samhällskunskap (Barton & Avery, 2016; Torrez & Claunch-Lebsack, 2014). Internationella studier har också visat att bedömningspraktiker kan se olika ut mellan olika stadier i skolan samtidigt som det kan finnas skillnader mellan länder (Brookhart, 2013; Kippers m.fl., 2018). I denna avhandling studeras bedömning i samhällskunskap på gymnasiet, ett ämne som läses på alla gymnasieprogram. Då lärares bedömningspraktik formas i en social och kulturell kontext (Allal, 2013) kan det vara av särskilt intresse att studera bedömningspraktiker bland svenska gymnasielärare i samhällskunskap, särskilt som detta ämne anses vara öppet och flexibelt, vilket ger lärarna stort handlingsutrymme i arbetet med bedömningar (Öberg & Bäckström, 2021).

Bedömning är något som alla som har gått i skolan har erfarenhet av. Dessa erfarenheter kan väcka olika känslor: en del kan minnas pressen inför prov, en upplevd orättvis bedömning, eller en inspirerande och rolig uppgift. Forskning har visat att elever upplever att bedömning är konstant närvarande i skolan (Hirsh, 2020; Vogt, 2017) och att bedömningar i skolan har betydelse för elevers motivation för skolarbete, liksom för deras självförtroende och känslor (Brookhart & Durkin, 2003; Löfgren m.fl., 2019; Pekrun m.fl., 2023). Elever kan känna sig mer eller mindre motiverade beroende på vilken typ av uppgift som ska utföras. De kan ha positiva eller negativa känslor inför att göra en uppgift. Dessa känslor, liksom elevens självförtroende, kan påverka deras prestation, samtidigt som känslor och självförtroende i sin tur kan ha påverkats av tidigare framgångar eller misslyckanden vid bedömningar.

Att känslor, självförtroende och motivation påverkas av bedömningar är inte så konstigt då bedömningar i skolan har stor betydelse för eleverna, särskilt när de också börjar få betyg. Dessa betyg används som urvalsgrund vid sökande av utbildningar och ibland också vid anställningar och har därför stor betydelse för elevernas framtid. I Sverige görs urval främst utifrån de betyg som lärarna sätter, vilket är relativt ovanligt i en internationell jämförelse (Lundahl m.fl., 2015, 2016). Dessa betyg grundas främst på uppgifter som lärarna själva konstruerar och många av lärarnas uppgifter blir därför *high-stakes*, det vill säga de har stor betydelse, för eleverna. Vad som bedöms och vilka metoder som används får därmed konsekvenser för eleverna. Forskning har visat att det kan påverka elevernas syn på vad som är viktigt att lära och därmed deras lärstrategier (Black, 2015; Brookhart & Durkin, 2003; Löfgren m.fl., 2019; Olovsson, 2015).

Eftersom betyg används som urvalsgrund, blir också frågan om likvärdig och rättvis bedömning viktig (Jönsson, 2023b). Svenska lärare har stora möjligheter att själva välja både innehåll och metoder för bedömning. Därför är det intressant med kunskap om lärares bedömningspraktiker för att på ett mer kvalificerat sätt kunna resonera om likvärdig och rättvis bedömning. För att bedömningar ska bli likvärdiga behöver de också grundas på relevanta underlag. Därmed blir också frågan om validitet viktig. Validitet handlar om att lärares bedömningar ska vara relevanta i relation till syftet med dem (Bonner, 2013; Kane, 2016; Messick, 1989). Oavsett om syftet är formativt eller summativt är validitet av vikt, då lärare behöver ändamålsenliga underlag såväl för att kunna ge relevant återkoppling som för att göra relevanta anpassningar av undervisningen eller för att sätta betyg (Dolin m.fl., 2018; Kane & Wools, 2020).

Bedömning är således en central arbetsuppgift för lärare. Vidare har det både internationellt och i Sverige blivit ett allt större intresse för bedömning i skolan, inte minst politiskt (Pettersson & Wester, 2020). Internationella kunskapsmätningar som PISA och TIMSS har satt fokus på utbildning och på olika utbildningssystems resultat. Detta har i Sverige bidragit till både reformer av och särskilda satsningar inom skolväsendet. Flera reformer har handlat om bedömning, särskilt betygsreformen 2011, som sedan även reviderats och fortfarande (år 2024) är under förändring (Betygsutredningen, 2020). Andra politiska beslut handlar om betyg i tidigare årskurser och fler nationella prov. Ett viktigt syfte med dessa reformer är att öka likvärdigheten i lärares bedömningar och betygsättning, då forskning och rapporter (Korp, 2006; Skolinspektionen, 2021; Skolverket, 2020) visat på problem med likvärdigheten mellan skolor och lärare; betyg sätts inte alltid på samma grunder. Detta kan bero på att lärare väger in annat som engagemang och flit. Det kan

också bero på att lärare ibland inte använder ändamålsenliga uppgifter som grund för betyg (Black m.fl., 2010). Genom att synliggöra lärares bedömningspraktik kan studien bidra till att utveckla lärares bedömningsuppgifter, vilket i sin tur kan bidra till en högre likvärdighet. Ovannämnda reformer har också inneburit ändrade villkor för lärares arbete med bedömningar (Wermke & Salokangas, 2021). Hur dessa nya villkor faktiskt påverkar lärares bedömningspraktik är en empirisk fråga.

Sammanfattningsvis kan konstateras att bedömningar är en viktig del av lärarprofessionen och skolan, samt att den påverkar individers möjligheter i såväl den fortsatta undervisningen som i framtida livsval. Det finns dock relativt få studier om lärares bedömningspraktik när det gäller vad som bedöms och hur, grunderna för lärares beslut om bedömningar samt vad som kan påverka dessa beslut. Den forskning som finns om lärares bedömningspraktik baseras ofta på lärares egna självbedömningar i form av enkäter (McMillan, 2013). Denna avhandling baseras på samhällskunskapslärares egna bedömningsuppgifter samt intervjuer, vilket bidrar till en rikare bild av lärares bedömningspraktik och grunderna för denna. Därmed kan en central del av lärarprofessionens uppgifter synliggöras.

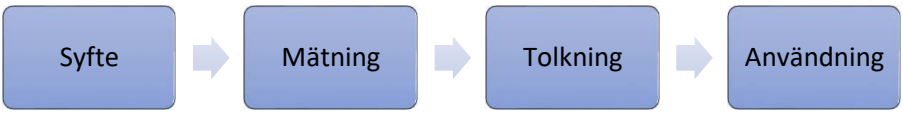
Syfte, frågeställningar och avgränsningar

Det övergripande syftet för denna avhandling är att undersöka samhällskunskapslärares bedömningspraktiker på gymnasiet, vad gäller syften, innehåll, förmågor, och metoder, samt sådant som kan påverka dessa praktiker. För att besvara syftet används följande forskningsfrågor:

1. Vilka syften uttrycker samhällskunskapslärares med sina bedömningar?
2. Vilket innehåll och vilka förmågor testar samhällskunskapslärare och vilka metoder använder de för att mäta detta?
3. Vilka samband mellan bedömningars syften, innehåll, förmågor och metoder uttrycker samhällskunskapslärare och hur relaterar det till validitet?
4. Hur resonerar samhällskunskapslärare om de faktorer som har betydelse för deras beslut gällande bedömningar och vilken agens kommer till uttryck?

”Resonera” i fråga fyra syftar på innebörden att framföra sin tankegång och redogöra för sina grunder för en uppfattning eller ett handlande (Svenska Akademien, 2023). I denna avhandling innebär resonera samhällskunskapslärarnas tankegångar om grunderna för deras beslut i bedömningsprocessen, både vilka grunderna (faktorerna) är och vilken betydelse som lärarna tillmäter dessa. Den första frågan besvaras främst i delstudie II, men berörs även i delstudie I. Fråga 2 och 3 behandlas

Figur 1. *Komponenter i en bedömningsprocess (McMillan 2018).*



främst i delstudierna I och III, medan fråga 4 behandlas i alla tre delstudierna.

I inledningen definieras bedömning som att den handlar om att samla in, tolka och använda information för att stödja lärarens beslut (McMillan, 2018). Det kan vara beslut inför planering av undervisning, under pågående undervisning, eller efter ett undervisningsområde, till exempel som underlag för betyg. Bedömning är en process, vilken kan beskrivas som bestående av komponenterna: syfte, mätning, tolkning samt användning (figur 1). Syftet med bedömningen (till exempel diagnostisera elevernas nuvarande kunskapsnivå, ge återkoppling eller underlag för betyg) bör, enligt modellen, påverka de andra komponenterna. Syftet bör styra val av lämpliga metoder för att mäta de kunskaper och förmågor som avses mätas. Sedan påverkar syftet tolkningen av kvaliteten i elevernas prestationer i relation till dessa kunskaper och förmågor, till exempel vilka bedömningskriterier som ska användas. Till sist bör syftet styra hur resultaten ska användas, exempelvis för beslut om fortsatt undervisning, återkoppling till eleverna eller som underlag för betyg. Lärares bedömningspraktik består av alla de beslut som lärare fattar i denna bedömningsprocess. I denna avhandling studeras dock bara de första komponenterna i bedömningsprocessen, det vill säga syfte och mätning, medan frågor om hur lärare tolkar och använder bedömningar inte behandlas.

I en annan modell över bedömningsprocessen lyfts också bedömningars konsekvenser fram (Mavrommatis, 1997), vilket skulle kunna ses som en femte ruta i modellen ovan. Vilka syften, vilket innehåll och vilka metoder som används kan få olika konsekvenser för elevernas lärande och reaktioner. I diskussionen diskuteras bedömningspraktikers konsekvenser genom att relatera studiens resultat till tidigare forskning om konsekvenser.

I avhandlingen används, utifrån ovanstående, fortsättningsvis begreppet bedömningspraktik i betydelsen av lärares beslut om syften, vilket innehåll och vilka förmågor som ska testas samt med vilka metoder.

2 Bedömningar förr och nu

Avhandlingen handlar om samhällskunskapslärares bedömningspraktik på gymnasiet. I detta avsnitt redovisas kontexten och den historiska bakgrunden till denna. Först beskrivs kort hur bedömningar utvecklats och använts genom historien, med fokus på Sverige. Därefter beskrivs hur kunskapssynen har förändrats med koppling till skola och bedömning. Sedan beskrivs det svenska betygssystemet och de styrdokument som berör bedömning inom gymnasieskolan. Till sist beskrivs ämnet samhällskunskap på gymnasiet.

Bedömningspraktiker genom historien

Det finns flera historiska spår som fortfarande influerar bedömningspraktiker i Sverige. Kunskapsbedömningar används för meritering, då de används för urval i olika sammanhang, ofta på grundval av betyg. Vidare har bedömningar länge präglats av skriftliga examinationer som fokuserat på utantillkunskaper, vilket det finns spår av än idag. En återkommande problematik har gällt brister med likvärdig bedömning. Nedan beskrivs dessa fyra aspekters historia: urval och betyg, metoder, vilka kunskaper som bedöms, samt likvärdighet.

Bedömningar användes för urval redan för åtminstone 2000–3000 år sedan i Kina (Lundahl, 2020). Även betyg har förekommit länge, till exempel användes betygen godkänt och underkänt ibland i de tidiga medeltida universiteten. Betyg användes i husförhören och sedan också i folkskolan som infördes 1842. Avgångsbetyg infördes 1877, vilka vid sekelskiftet blev ett krav för att få anställas i bland annat industrin. I och med detta formaliserades och standardiserades folkskolans betyg. Elever med godkända avgångsbetyg tilläts därefter att göra inträdesprov till läroverken. Läroverken examinerade också sina elever samtidigt som universiteten hade inträdesprov fram till 1862. då de avskaffades. I stället infördes censorer från universiteten som också skulle bedöma vid läroverken för utfärdande av avgångsexamen, vilken år 1906 bytte namn till studentexamen. Bedömningarna grundades både på lärarnas skriftliga prövningar och censorernas muntliga förhör. Denna ordning fanns kvar in på 1960-talet, varefter enbart betyg kom att användas.

I samband med grundskolereformerna på 1960-talet infördes normrelaterade betyg, där elevernas betyg skulle sättas i relation till alla andra elever i riket (Samuelsson, 2017).. Detta betygsystems främsta funktion var att sortera och användas för urval. En omfattande kritik växte med tiden fram, dels mot att betygen användes fel och att de högre betygen kunde ”ta slut”, dels för att de inte visade vilka kunskaper eleverna hade.

Kritiken ledde till införandet av ett kriterierelaterat betygssystem 1994. Detta innebar att betygen skulle sättas i relation till fastställda betygskriterier. Denna förändring innebar också en minskad tonvikt på bedömningarnas sorterande funktion och i stället betonades bedömning för lärande (Lundahl, 2020). I praktiken fortsatte dock betygen att användas för urval.

Historiskt har ofta utantillkunskaper premierats (Lundahl, 2020). Så var det både i läroverken som växte fram i Sverige under 1500-talet och i folkskolan. Dessa kunskaper testades ofta i muntliga förhör och skriftliga examinationer. I läroplanerna för grundskolan, och sedan även för gymnasiet, började dock även andra kunskaper betonas såsom kritiskt tänkande och problemlösning. De normrelaterade betygen skulle således grundas på fler förmågor, men i praktiken kom många gånger faktaorienterade prov att dominera (Mattson, 1989). Läroplanerna från 1994 innebar en tydligare övergång från den faktaorienterade skolan till en betoning på förmågor (Lundahl, 2020). Detta uttrycktes i läroplanens kunskapssyn där kunskap ses som en helhet av kunskapsformerna fakta, förståelse, färdighet och förtrogenhet, vilken är densamma i läroplanerna från 2011 och 2022 (Carlgren m.fl., 2009). Det tycks dock som att många lärare har fortsatt att betona faktakunskaper, även efter reformerna 2011 (Gomez, 2017; Hartell & Strimel, 2019; Kakoulidou, 2021; Rosenlund, 2016; Svingby, 1998).

Likvärdighet har också varit problematiskt länge. Redan under folkskolan utvecklade lärare olika betygspraktiker, vilket ledde till att folkskoleinspektörer infördes, mallar utarbetades, samt försök med standardiserade tester (Lundahl, 2020). I Sverige utvecklades en bedömningskultur som skiljer sig från övriga Norden (Tveit, 2018). Till en början var det en liknande utveckling i slutet av 1800-talet fram till 1940-talet i och med införande av folkskolor, meritering och betyg. Sverige kom dock att influeras mer av en amerikansk psykometrisk tradition med fokus på produkt än till exempel Norge som mer präglats av en europeisk tradition med fokus på process. Det har i Sverige sedan 1930-talet inneburit en större tilltro till nationella tester för att kontrollera och styra lärares bedömningar. Lärarna i Norge har däremot bibehållit tilltro och ansvar för bedömningar och nationella tester har inte fått samma genomslag, i vart fall inte förrän senare under 2000-talet. Den psykometriska traditionen, tillsammans med friskolereformen på 1990-talet, bidrog till att en performativ diskurs, vilken betonar mål och resultat (exempelvis betyg), snabbare fick genomslag i Sverige och att nationella prov har fått större betydelse utan att möta något större motstånd.

Särskilt betygsreformen 2011 kom att innebära en stark performativ diskurs, med en betoning på summativ bedömning för betygsättning (Strandler, 2017; Wahlström & Sundberg, 2015). I ett kriterierelaterat

betygssystem ses lärarna mer som utförare av de reformer som genomförs, än som medskapare av utbildningen. Wahlström och Sundberg menar att detta tillsammans med betoningen på centralt innehåll och framför allt på kunskapskraven i styrdokumenten riskerar att leda till en instrumentalisering av lärares undervisning, vilket också riskerar att begränsa lärares handlingsutrymme.

Skolans styrdokument (som inkluderar läroplanen, ämnes-/kursplaner, allmänna råd och andra kommentarmaterial från Skolverket) är enligt en enkätstudie med 573 grundskole- och gymnasielärare det som främst inramar bedömningsarbetet (Wermke & Salokangas, 2021). Dessa får därför stor betydelse för lärares bedömningspraktik och dessa dokument definierar också i hög grad hur lärarna ser på sig själva som tjänstemän. Styrdokumenten är restriktiva, men de ger de också lärarna en känsla av säkerhet. Samtidigt lyfter svenska ämneslärare fram akademisk kunskap som en viktig faktor och en akademisk inramning skulle kunna skydda en professionell autonomi. Reformerna innebar minskad autonomi för lärare och en avprofessionalisering, vilket förstärks av marknadiseringen av det svenska skolsystemet, där betyg blir en konkurrensfaktor (Wermke & Salokangas, 2021). Skolors och lärares betygsättning granskas av både skoladministratörer och föräldrar, vilket också bidrar till en ökad administration då lärarna behöver dokumentera mer för att kunna rättfärdiga sin betygsättning. Samtidigt har lärarna kvar makten över att sätta betyg, även om de är utsatta för kontroll.

Sammanfattningsvis finns en lång historia där bedömningar och betyg används för urval vid både kommande studier och anställning. Detta gör lärares bedömningar viktiga för eleverna, eftersom de påverkar deras framtida möjligheter. Vidare har lärares bedömningar dominerats av skriftliga prov, vilka främst testat faktakunskaper. Oavsett betygssystem har det också funnits problem med likvärdigheten, vilket i Sverige har hanterats med kontroll genom nationella prov. De senaste betygsreformerna har inneburit en performativ diskurs som i den svenska testningskulturen fått stort genomslag i lärares bedömningspraktiker, vilka begränsas av denna diskurs. Samtidigt är det i slutändan lärarna som utformar sina bedömningar och sätter betyg och det finns ett utrymme för en professionell autonomi.

Förändrad syn på bedömning

Synen på bedömning har förändrats, både internationellt och i Sverige under senare delen av 1900-talet (Lindström, 2013). Forskning och litteratur om bedömning pekar på en förskjutning av perspektiv från kontroll av kunskaper till att främja lärande, där lärares återkoppling med fokus på det fortsatta lärandet till eleven blir central. Ofta kallas detta

formativ bedömning. Denna bygger på en förändrad kunskapssyn där förståelse, färdigheter och förmågor betonas och faktakunskaper tonas ned. Den innebär också större fokus på processer än på produkter. Denna förändring har också påverkat de svenska läroplanerna och betygssystemet med kunskapssynen i läroplanerna 1994 som betonar olika kunskapsformer och som även utgör grunden i läroplanerna från 2011 och 2022.

Formativ bedömning innebär att bedömningens fokus är på att bedömningar används för att stödja och utveckla elevens lärande (Cizek, 2009a; Lindberg & Hirsh, 2015; Roos & Hamilton, 2005). Denna utvecklades med rötter i 1960-talet, utifrån bland annat konstruktivistiska teorier, och har sedan spritts, framför allt från slutet av 1990-talet och framåt. Forskningen om formativ bedömning har ökat markant under 2000-talet och än mer under 2010-talet, även i Sverige (Lindberg & Hirsh, 2015). Formativ bedömning har under 2010-talet fortsatt att spridas och enligt Wahlström och Sundberg (2015) råder en stark formativ diskurs, vilken även finns i läroplanens kunskapssyn. Skolverket har också alltmer lyft fram detta perspektiv genom sina kommentarmaterial under 2010-talet och framåt (Skolverket, 2011b, 2012, 2018, 2021, 2022a).

I sitt arbete med bedömningar befinner sig således lärare i denna kontext av både en stark formativ diskurs och en summativ och performativ diskurs (Strandler, 2017; Wahlström & Sundberg, 2015). Det är i detta spänningsfält som deras bedömningspraktik formas.

Det svenska betygssystemet

Betygsreformen 1994 innebar en stor förändring då Sverige bytte från normrelaterade till kriterierelaterade betyg, vilket är det system som fortfarande används. Senare reformer har inte inneburit något nytt betygssystem, utan har varit ändringar i utformningen av det kriterierelaterade systemet.

Det normrelaterade betygssystemet, som infördes 1962 i grundskolan och 1966 på gymnasiet, innebar att elevernas prestationer jämfördes med en norm, vilken var alla andra elevers (i årskullen) medelprestation (Betygsberedningen, 2008). Betygen skulle fördelas enligt en normalfördelning[2]. Det fanns inga underkända betyg i detta system, men streck kunde ges för elever med hög frånvaro. Tanken var att jämförelsen skulle göras mot riket och som stöd infördes standardprov i några ämnen på grundskolan och centrala prov på gymnasiet. I praktiken kom dock många lärare att jämföra och fördela betygen inom respektive klass.

[2] Normalfördelningskurvan innebar att 7% av eleverna i riket skulle få betyget 1, 24% betyget 2, 38% betyget 3, 24% betyget 4 samt 7% betyget 5.

Det fanns en omfattande kritik mot systemet. Bland annat ansågs systemet leda till konkurrens, vilket stod i motsats till läroplanens mål om samarbete (Betygsberedningen, 2008). Vidare kritiserades att betygen inte sa något om elevens kunskaper, utan endast hur deras kunskaper stod sig i relation till andra. Styrkan med systemet var dock att den fungerade väl för urval, vilket också var främsta skälet till att det infördes. Hinder för reformer var också att det befarades vara svårt att utforma tydliga och användbara kriterier.

Den omfattande kritiken ledde dock till reformerna 1994 då de kriterierelaterade betygen infördes (Betygsberedningen, 2008). Dessa innebar att elevernas prestationer skulle jämföras med formulerade betygskriterier på olika nivåer, i skalan G–VG–MVG, där det på gymnasiet också fanns ett underkänt betyg, medan det i grundskolan inte sattes något betyg om eleven inte uppnådde kriterierna för godkänt. Delstudie I i denna avhandling genomfördes under detta betygssystem.

Betygsberedningen arbetade med betygsreformen parallellt med läroplanskommittén, vilket ledde till att läroplanens kunskapssyn med fakta, förståelse, färdighet, och förtrogenhet inte helt slog igenom i utformandet av betygskriterierna (Betygsberedningen, 2008). Läroplanens syn är att de fyra kunskapsformerna inte är hierarkiska utan parallella och förutsättande varandra. Betygsutredningen hade däremot en mer taxonomisk syn och betygskriterierna kom att ha en hierarkisk struktur, där godkända betyg ofta främst innebar faktakunskaper medan högre betyg innebar olika grader av förståelse, resonerande och analytiska färdigheter samt förtrogenhet.

Med tiden visades en del brister även med det nya betygssystemet (Betygsberedningen, 2008). Bland annat fanns kritik mot de hierarkiska betygskriterierna. Den främsta kritiken handlade dock om bristande likvärdighet och rättvisa, vilket också bidrog till en omfattande betygsinflation. Otydliga kriterier ansågs vara en viktig orsak, liksom att lärare vid betygsättningen skulle använda en kompensatorisk princip, vilken innebar att en elevs styrkor i relation till vissa kriterier, kunde väga upp brister i relation till andra kriterier. En annan kritik var att det var för få betygssteg och många menade att det var för stor bredd inom betygen, till exempel mellan de som precis uppnått Godkänt och de som hamnat strax under Väl Godkänt.

I betygsreformen 2011 ändrades således utformningen av betygskriterierna, vilka nu benämndes kunskapskrav för att markera att kompensationsmodellen ersatts med en besiktningsmodell där samtliga kriterier för ett betyg behövde vara uppnådda för att det betyget skulle sättas (Betygsberedningen, 2008; Betygsutredningen, 2020). Styrkor kunde inte längre kompensera brister. Betygsskalan ändrades och utökades till A–F. Betygen A–E var godkända betyg, med A som högsta betyg och F

innebar ett underkänt betyg (som nu även tillämpades i grundskolan). Kunskapskrav formulerades för betygen A, C och E, medan B och D skulle ges till dem som uppfyllt samtliga kunskapskrav för betygsnivån under och till övervägande del uppfyllt kunskapskraven för det högre betyget. Kunskapskraven var inte längre hierarkiska med en progression i olika kunskapsformer. I stället fanns samma förmågor på alla betygsnivåerna men med en progression, vilken uttrycktes med olika värdeord. Därmed kom läroplanens icke-hierarkiska kunskapssyn att prägla även kunskapskravens utformning. Ett exempel på progressionen är ett kriterium för betyget E i samhällskunskap (Skolverket, 2022c):

> Eleven kan **översiktligt** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna.

Progressionen uttrycks genom att det fetstilta värdeordet byts ut mot **utförligt** för C, respektive **utförligt och nyanserat** för A. Delstudierna II och III gjordes när detta betygssystem gällde.

Även detta system blev kritiserat. Den främsta kritiken riktades mot besiktningsmodellen, då effekten av denna blev att elevernas betyg kom att avgöras av deras sämsta prestation (Betygsutredningen, 2020). Det fanns också kritik mot kunskapskravens formuleringar, vilka ansågs krångliga och svårtolkade. Detta ledde till ytterligare en reform som trädde i kraft 2022. Skalan A–F bibehölls och det är fortfarande för betygen A, C och E som det finns kriterier formulerade. Förändringen innebar främst en förenkling av kriterierna och för grundskolan en större betoning på faktakunskaper, särskilt i de lägre årskurserna. För betyget E gäller fortfarande besiktningsmodellen, det vill säga alla kriterier måste vara uppfyllda, men för betygen C och A återinfördes kompensationsmodellen, där styrkor kan uppväga svagheter. Läraren ska göra en sammantagen bedömning, vilket innebär att sätta det betyg som bäst motsvarar elevens kunskaper. Om läraren bedömer att en elevs kunskaper ligger mellan E och C eller mellan C och A sätts betyget D respektive B. Kriterierna benämns återigen betygskriterier för att markera att de inte längre är krav, utan de uttrycker kännetecken för respektive betyg.

Skolverket har också gett ut allmänna råd och kommentarmaterial om bedömning för att ge råd och riktlinjer om hur styrdokumenten ska tillämpas. I dessa har Skolverket också lyft fram det formativa perspektivet på bedömning. I de tidigare kommentarmaterialen (Skolverket, 2011b, 2012) används visserligen inte begreppet formativ bedömning men flera råd pekar i den riktningen, exempelvis talas om vikten av att stödja kunskapsutvecklingen och att ge framåtsyftande återkoppling. I senare material blir det än tydligare då det även explicit står om formativ

bedömning som också betonas (Skolverket, 2018, 2021, 2022a). I stället tonas summativ bedömning ned, till exempel att lärare varken måste eller bör ge betygsomdömen på enskilda uppgifter.

Sammanfattningsvis har bedömning i Sverige gått från normrelaterad till kriterierelaterad, vilket också har inneburit en förskjutning från att främst bedöma faktakunskaper till att också bedöma förståelse och resonerande förmågor. Vidare har synen på bedömning utvecklats i två riktningar, dels mot en stark performativ diskurs med betoning på summativ bedömning, främst betygsättning, dels mot en formativ diskurs som betonar användningen av bedömning för att stärka elevernas lärande. Lärare behöver ta hänsyn till båda dessa olika aspekter när de utformar sina bedömningar.

Ämnet samhällskunskap på gymnasiet

De flesta elever läser efter grundskolan vidare på gymnasiet där de går antingen ett yrkesprogram[3] eller ett högskoleförberedande program (Skolverket, 2011a). Utbildningen läses i ett antal olika kurser (vilka totalt omfattar 2500 poäng), varav några är så kallade gymnasiegemensamma ämnen, vilka läses oavsett vilket program eleven går. Ett av dessa ämnen är samhällskunskap där den första kursen är obligatorisk. Den finns i två versioner, Samhällskunskap 1a1 (se bilaga 1), vilken läses på yrkesprogrammen, och Samhällskunskap 1b (se bilaga 2), vilken läses på de högskoleförberedande programmen. Samhällskunskap 1a1 är hälften så omfattande (50 poäng) jämfört med 1b (100 poäng). Elever på yrkesprogrammen kan sedan läsa till kursen 1a2 för att komplettera motsvarande 1b-kursen. Dessa kurser är grundläggande kurser med ett brett innehåll, vilka sedan kan fördjupas i andra samhällskunskapskurser (Samhällskunskap 2, Samhällskunskap 3, Internationella relationer, samt Internationell ekonomi). I denna avhandling fokuseras bedömning i den obligatoriska första kursen Samhällskunskap 1a1/1b[4].

Alla kurser har en gemensam struktur som består av ämnets syfte, centralt innehåll och kursens betygskriterier (Skolverket, 2011a). Det centrala innehållet kan vara både faktainnehåll och förmågor som ska utvecklas. Det centrala innehållet i samhällskunskap 1b innefattar följande (Skolverket, 2022c):

- Demokrati och politiska system på lokal, nationell och EU-nivå, samt politiska ideologier.

[3] Före 2011 hette det yrkesförberedande program.

[4] I delstudie I, som genomfördes före reformen 2011, motsvaras detta av kursen Samhällskunskap A. Då fanns det bara en första obligatorisk samhällskunskapskurs, vilken var samma för både yrkes- och studieförberedande program. Den kursen beskrivs mer utförligt i delstudie I.

- Mänskliga rättigheter.
- Folkrätten i väpnade konflikter.
- Arbetsmarknaden och dess parter, arbetsrätt och arbetsmiljö.
- Gruppers och individers identitet, relationer och levnadsvillkor.
- Samhällsekonomi.
- Privatekonomi och konsumenträtt.
- Media och informationsteknologi.
- Källkritik.
- Samhällsvetenskapliga begrepp, teorier och metoder.
- Presentation i olika former (både muntligt och skriftligt).

Samhällskunskap 1a1 har i stort sett samma centrala innehåll, men saknar politiska ideologier, samhällsekonomi, samhällsvetenskapliga metoder, samt har mindre om medier.

I samhällskunskap 1a1/1b är betygskriterierna, liksom i de flesta kurser, mer generella i relation till det centrala innehållet (Skolverket, 2022c). Det finns två betygskriterier som specifikt pekar ut ett centralt innehåll som måste bedömas; att eleverna måste kunna redogöra för de mänskliga rättigheterna respektive för frågor om privatekonomi och konsumenträtt.

Ämnet samhällskunskap anses vara öppet och flexibelt med en svag inramning, och anses sakna en tydlig ämneskanon (Hadley & Young, 2018; Linde, 2021; Öberg & Bäckström, 2021). Denna syn har dock kritiserats, då detta främst grundats på läroplansstudier (Olsson, 2016), medan empiriska studier pekar på ett tydligare ämne med en kärna bestående av statsvetenskap och samhällsekonomi (Odenstad, 2011; Olsson, 2016; Öberg & Bäckström, 2021). Även om ämnes- och kursplanen kan vara vag, tycks det således finnas en innehållslig kärna i lärares undervisning och bedömning, vilket indikerar att det finns en kanonisk ämnestradition (Öberg & Bäckström, 2021). Samtidigt finns det en stor variation i lärares undervisnings- och bedömningsmetoder, vilket Öberg och Bäckström förklarar med lärares olika preferenser.

Ämnet samhällskunskap har en karaktär inriktad mot medborgerlig fostran (Englund, 1986; Torrez & Claunch-Lebsack, 2014), med fokus på medborgarskap, resonerande förmågor och kritiskt tänkande (Arevik & Hartzell, 2015; Möllenborg, 2023; Sandahl m.fl., 2022; Svingby, 1989; Wahlström m.fl., 2020). Möllenborg beskriver samhällskunskap som ett demokrati-, ekonomi-, analys- och funktionellt ämne (allmänbildande). Hennes studie visade också att undervisning tenderar att skilja sig mellan högskoleförberedande program och yrkesprogram. Wahlström m.fl. (2020) menar att ämnets kursplan involverar prestationsbaserade generiska kompetenser, vilka de menar är mer betonade än ämnesbaserade förmågor och innehåll. Samtidigt menar de att dessa utgör starka ramar

utifrån en performativ diskurs som betonar betygskriterierna och tvingar lärarna att samla in underlag för betyg. Denna bild av fokus på generiska kompetenser motsägs dock av Möllenborgs (2023) studie liksom i denna avhandling av lärarna i delstudie II och III, vilka uttrycker dessa generiska kompetenser i relation till ett ämnesinnehåll och därför tycks uppfatta generella förmågor som att analysera som annorlunda och specifika för samhällskunskap jämfört med andra ämnen (Jansson & Löfgren, 2022).

I denna avhandling studeras lärares bedömningspraktik i samhällskunskap på gymnasiet, där ämnets styrdokument, karaktär och tradition är aspekter som kan ha betydelse för lärares syften och beslut när de utformar sina bedömningar.

3 Teoretisk bakgrund

I avhandlingen studeras vad lärare testar i sina bedömningsuppgifter, vilket också innefattar olika former av kunskaper. Som stöd för analyserna av kunskapsformer användes Blooms reviderade taxonomi i delstudie I och III för att analysera lärares bedömningsuppgifter, kompletterat med intervjuerna. Vidare relaterar en av forskningsfrågorna till validitet. Taxonomin och validitet redovisas därför nedan.

En del av syftet handlar om lärares resonemang om olika faktorer som har betydelse för deras bedömningspraktik. Under arbetet med delstudierna och avhandlingen har det varit en process i valet och användningen av teori, där olika teorier och modeller har använts för att bidra till förståelsen av lärares bedömningspraktik. I den första delstudien användes ramfaktorteori för att analysera olika faktorer som inramar lärares arbete och diskutera lärares frirum. I delstudie II användes en modell om interna och externa faktorer i lärares beslutsprocess om bedömningar i diskussionen om lärares syften med bedömningar. I delstudie III användes ett ekologiskt perspektiv på läraragens (*teacher agency*), vilken också används i avhandlingens diskussion.

Dessa teorier har en del likheter, men också skillnader som kompletterar varandra. Alla tre teorierna belyser olika faktorer som inramar och påverkar lärares arbete med bedömningar, men de använder olika begrepp. Det finns också en utvidgning mellan dem i vilka faktorer som finns med. Ett gemensamt drag mellan dem är perspektivet att det finns faktorer som begränsar och möjliggör lärares handlingar. I ramfaktorteorin kallas detta ramfaktorer, medan det i läraragens motsvaras av olika kontextuella villkor. I beslutsmodellen relaterar detta till interna och externa faktorer, även om dessa beskrivs mer som påverkande aspekter än som ramar. Ett viktigt begrepp i ramfaktorteorin är frirum, vilket påminner om agens, men agens är ett vidare begrepp. Dessa teorier och sambanden mellan dem redovisas nedan.

Blooms reviderade taxonomi

I analyserna användes Blooms reviderade taxonomi för att kategorisera vilka olika kunskapsformer som testas i lärares olika bedömningsuppgifter. Denna taxonomi har utvärderats i en studie av Näsström (2008) som kom fram till att taxonomin är användbar för att analysera kunskapsformer. Den har bland annat använts i svenska studier om bedömning i historia för att analysera vilka kunskapsformer som bedöms (Kakoulidou, 2021; Rosenlund, 2016).

***Figur 2**. Blooms reviderade taxonomi (Andersson m.fl., 2001).*

Kunskapsdimensionen	Den kognitiva dimensionen						
	Blooms reviderade taxonomi	Minnas	Förstå	Tillämpa	Analysera	Värdera	Skapa
	Faktakunskap						
	Konceptuell kunskap						
	Procedur-kunskap						
	Metakognitiv kunskap						

Blooms reviderade taxonomi är tvådimensionell (figur 2) (Anderson m.fl., 2001). Det innebär att taxonomin skiljer på innehållet, *kunskapsdimensionen*, och på med vilken förmåga detta innehåll kan hanteras, *den kognitiva dimensionen*. Faktakunskap i kunskapsdimensionen kan till exempel kognitivt både kommas ihåg, förstås eller användas i en analys. Kunskapsdimensionen består av faktakunskaper, konceptuell kunskap[5], procedurkunskap samt metakognitiv kunskap. Den kognitiva dimensionen delas in i minnas, förstå, tillämpa, analysera, värdera och skapa. En mer detaljerad beskrivning finns i delstudie I.

I delstudie I användes denna taxonomi för att kategorisera de olika frågorna i lärarnas prov. I delstudie III användes taxonomin i en förenklad form för att kategorisera vilka kunskaper som testas i lärarnas olika inskickade bedömningsuppgifter.

Validitet

Två av avhandlingens frågor handlar om lärarnas syften med bedömningar, respektive om samband mellan syften, testade kunskaper och

[5] I delstudie I benämns denna begreppskunskap i enlighet med en översättning från Institutionen för beteendevetenskapliga mätningar, Umeå Universitet. Uttrycket begreppskunskap riskerar dock att sammanblandas med kunskaper om terminologi, vilket är en del av faktakunskap. Den mer direkta översättningen konceptuell kunskap, från engelskans *conceptual knowledge*, signalerar tydligare att det handlar om en mer abstrakt och generell nivå.

förmågor och metoder. Därmed är frågan om validitet också intressant för denna avhandling. Validitet i bedömningar har länge ansetts vara viktigt, då bedömningar, för att de ska anses vara av hög kvalitet, behöver vara relevanta utifrån syftet med dem (McMillan, 2018). Det råder inte konsensus om hur validitet ska definieras, men det viktiga är att vara tydlig med vilken definition som används i den specifika kontexten (Markus, 2016).

En vanlig definition är att validitet innebär att tolkningar och användning av bedömningar ska vara rimliga (Messick, 1989; Newton & Shaw, 2014; Sireci, 2016). Flera aspekter av validitet brukar lyftas fram: innehållsvaliditet, kriterievaliditet (vilken inkluderar samtidig respektive prediktiv validitet), begreppsvaliditet samt konsekvensvaliditet. Innehållsvaliditet handlar om i vilken grad bedömningen är representativ för området som ska bedömas. Kriterievaliditet handlar om hur väl det som testas korrelerar med ett kriterium som definierats i förväg. Detta kan antingen vara samtidig validitet, vilket innebär att bedömningen ger samma resultat som andra bedömningar av samma kriterium, eller prediktiv validitet som handlar om hur väl bedömningen kan förutsäga framtida prestationer. Begreppsvaliditet handlar om hur väl en bedömning mäter ett abstrakt begrepp, till exempel intelligens. I skolans miljö skulle det kunna handla om förmågor som att förstå eller att analysera. Konsekvensvaliditet handlar om i vilken grad en bedömning leder till rimliga utfall, både förväntade och oväntade.

I denna avhandlingskappa används Messicks (1989) definition av validitet. Han förenar validitet i en universell definition där begreppsvaliditet utgör ett paraply för alla de andra ovan nämnda aspekterna av validitet. Enligt Messick handlar validitet om i vilken grad empirisk evidens stödjer tolkningar och användning av bedömningars resultat. Han har utvecklat en modell (figur 3) med olika aspekter av validitet i relation till tolkning och användning. En aspekt handlar om evidensen för tolkning och användning. Det innebär att underlagen behöver vara relevanta för det uppgiften avser bedöma, det vill säga i relation till vilket innehåll och vilka förmågor som ska bedömas. Validitet handlar om i vilken grad tolkningar och användning är rimliga och relevanta utifrån bedömningens resultat. Om till exempel syftet är att bedöma elevers analysförmåga,

Figur 3. *Aspekter av validitet (Messick, 1989, s. 20)*

	Tolkning av resultat	*Användning av resultat*
Evidensgrund	(Begrepps-)validitet	(Begrepps-)validitet/nytta
Konsekvenser	Påverkan på värderingar	Sociala konsekvenser

men uppgiften är ett faktaprov, går det inte att dra valida slutsatser om elevers analysförmåga (men däremot om deras faktakunskaper). Den andra aspekten handlar om bedömningars konsekvenser, dels om hur tolkningen av bedömningen kan påverka uppfattningar och värderingar, till exempel om vad som är viktig kunskap, dels om vilka sociala konsekvenser användningen av bedömningen kan få, till exempel för motivation och självkänsla. Validiteten kan dock aldrig helt bevisas, utan det handlar om att motivera tolkning och användning med stöd i evidens. Argumentet kan då vara mer eller mindre valitt.

Hur en bedömningsuppgift är konstruerad är således avgörande för validiteten. Messick (1989) beskriver olika hot mot validiteten utifrån hur en bedömning är utformad: underrepresentation (*construct underrepresentation*) och irrelevans (*construct-irrelevant variance*). Underrepresentation innebär att testet inte täcker hela det kunskapsområde som är tänkt att bedömas. En uppgift behöver ha ett representativt innehåll för det som ska mätas, men också mäta de förmågor som avses mätas. Validitet påverkas negativt när betyg innebär olika saker mellan olika skolor, ämnen, lärares klasser eller för olika typer av elever. Irrelevans innebär att uppgifter inte är relevanta för att mäta de kunskaper som ska bedömas. Det kan också handla om att andra irrelevanta faktorer kan påverka resultatet och bedömningen, till exempel olika former av bias, det vill säga faktorer som snedvrider bedömningen. Dessa kan finnas både i själva uppgiften och i tolkningen av resultaten. Det kan vara att uppgifter är konstruerade så att de gynnar eller missgynnar vissa elever, till exempel att en uppgift kräver språkliga färdigheter även om det inte är det språkliga som avses mätas, eller att lärarens uppfattning av en elev påverkar bedömningen. För att minska irrelevans är det viktigt att ha flera underlag för bedömningen. Messick skriver också om konvergent evidens, vilket innebär att tolkningen av en bedömning behöver stämma överens med andra mått av samma förmåga som testas. Det betyder att det också behövs en variation. Med andra ord är det viktigt att en uppgift är konstruerad så att den är representativ för både innehåll och de förmågor som ska bedömas, men också att den inte innehåller sådant som kan snedvrida resultaten eller försvåra tolkningen av elevernas prestationer.

Validitet handlar således om att ha ändamålsenliga underlag för att tolkningen och användningen ska bli valid. I avhandlingen studeras visserligen inte lärares tolkning och användning av bedömningar utan deras praktik gällande underlagen, vilket gör att validitet är relevant att diskutera i relation till studien. Validitet är inte bara av vikt vid summativ bedömning. Även vid formativ bedömning behöver läraren ändamålsenliga underlag i relation till undervisningens mål för att kunna ge relevant återkoppling och fatta rimliga beslut om fortsatt undervisning (Dolin

m.fl., 2018; Gu, 2021; Wellberg, 2023). I avhandlingens diskussion relateras därför studiens resultat till evidensgrunden i Messicks modell, eftersom denna har betydelse för validiteten i tolkning och användning.

Messicks modell är omdiskuterad och vissa menar att konsekvensvaliditet inte är relevant för validitet (Newton & Shaw, 2014). Samtidigt är forskare överens om att konsekvenser är viktiga att beakta, men kritiker menar att det inte är en del av validitet. I avhandlingens diskussion diskuteras också, med stöd av Messicks modell, konsekvenser av bedömningar i relation till både värderingar och sociala aspekter.

Ramfaktorteori

Ramfaktorteorin syftar till att förklara undervisningsprocesser i klassrummet utifrån olika faktorer (Lundgren, 1989). Eftersom bedömning är en central och integrerad del i lärares undervisning kan teorin även bidra till förståelsen av lärares bedömningspraktik. Teorin innebär att det finns ett samband mellan ramar av olika slag, undervisningsprocesser och resultat (Lundgren, 1989). Ramar förklaras som faktorer som möjliggör vissa handlingar i klassrummet, men begränsar andra. Ramar kan vara av olika slag: pedagogiska (skolans styrdokument), administrativa (som arbetslag, scheman och arbetstidsregler), resursrelaterade (exempelvis läromedel, antal undervisningstimmar och lokaler), organisationsrelaterade (skolans sociala förhållanden och rådande kultur) eller elevrelaterade (elevernas förmågor och bakgrund) (Imsen, 1999; Lundgren, 1989).

Ramfaktorteorin kan hjälpa till att förklara och förstå pedagogiska processer, men inte förutsäga dem (Gustafsson, 1999). Teorin kan förklara samband mellan ramar och undervisningsprocesser och hur resultat grundas i en viss social och historisk kontext (Lindblad m.fl., 1999). Ytterligare en begränsning med ramfaktorteorin är att den inte tydligt beaktar läraren som aktör (Lindblad & Sahlström, 1999). Detta kan bero på att teorin inte utvecklades för att förklara lärares villkor, utan för att utvärdera konsekvenser av politiska beslut (Lindblad m.fl., 1999). Visserligen tar teorin upp yttre och inre ramar, men inre ramar handlar om hur lärarna uppfattar de yttre ramarna, vilket påverkar lärarens upplevda frirum (Lindblad & Sahlström, 1999). De upplevda ramarna kan ofta vara snävare och frirummet kan vara större inom de reella ramarna och det kan även finnas ett frirum som lärarna inte använder (Gustafsson, 1999). Ramfaktorteorin används i delstudie I för att förstå hur olika ramfaktorer inramar lärares bedömningspraktik, samtidigt som lärare kan uppleva ett visst frirum och agera inom dessa ramar.

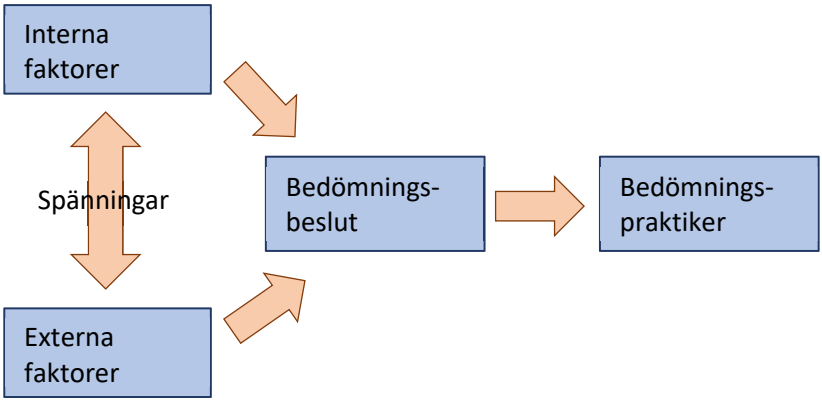
Undervisningen påverkas således också av de enskilda lärarna, till exempel av hur bred eller snäv undervisningsrepertoar de har (Lindblad

m.fl., 1999). Repertoarer handlar om lärarens bredd i arbetssätt och metoder (förmedlingsrepertoar), men också bredd (och djup) i stoffkunnandet (stoffrepertoar). Ju bredare repertoar, desto större frirum besitter en lärare i relation till olika ramar. Betydelsen av lärares egna intressen och preferenser beaktas inte i ramfaktorteorin, men detta finns med i beslutsmodellen som interna faktorer, vilka tillsammans med externa faktorer har betydelse för lärares beslut om bedömningar.

Interna och externa faktorer

Modellen om lärares bedömningsbeslut (figur 4) beskriver att lärares bedömningspraktik formas utifrån interna och externa faktorer (McMillan, 2003; McMillan & Workman, 1999). Interna faktorer handlar om lärarens egna kunskaper, uppfattningar, förväntningar och värderingar. Det handlar om syn på kunskap och lärande, en vilja att stötta elever till att lyckas, individuella anpassningar, liksom en vilja att öka elevernas motivation och gynna deras förståelse. Externa faktorer kommer utifrån och utgörs av styrdokument och policy, föräldrar, externa prov (till exempel nationella prov) samt att lärare utvärderas externt. Interna och externa faktorer har betydelse för lärarnas beslut om bedömningar, till exempel om vilka bedömningsformer som ska användas, utformning av bedömningsuppgifter och matchning mellan lärandemål och bedömning (McMillan 2018). Lärarnas beslutsprocesser om bedömning ligger till grund för deras bedömnings- och betygsättningspraktiker. Samtidigt finns det spänningar mellan dessa interna och externa faktorer och lärares bedömningspraktik formas i detta spänningsfält.

Figur 4. *En modell över lärares beslutsfattande om klassrumsbedömningar (modifierad från McMillan, 2003).*



Beslutsmodellen användes i delstudie II för att kategorisera lärares olika syften med bedömningar utifrån om de är interna eller externa. Detta bidrar till en förståelse för att lärares syften med bedömningar kan komma både inifrån lärare själva och utifrån, till exempel i form av styrdokument. Spänningar mellan olika syften blir också synliggjorda.

En brist med beslutsmodellen är dock att den inte har med så många externa faktorer. Till exempel saknas administrativa och materiella villkor som schema och arbetstid. Ramfaktorteorin bidrar således med fler påverkande faktorer än beslutsmodellen och dessa två kompletterar varandra.

Lärare är dock inte passiva offer inom ramarna eller de interna och externa faktorerna. De är också aktörer inom dessa (Lindblad m.fl., 1999). Aktörerna är rationella genom att de använder sitt praktiska förnuft i den historiska och sociala kontexten. Handlingarna blir logiska i relation till olika situationers möjligheter och begränsningar. De kan variera beroende på kontexten i stunden och de rådande förhållandena går också att påverka. Detta leder till en annan teori, nämligen om läraragens (*teacher agency*), vilken redovisas nedan.

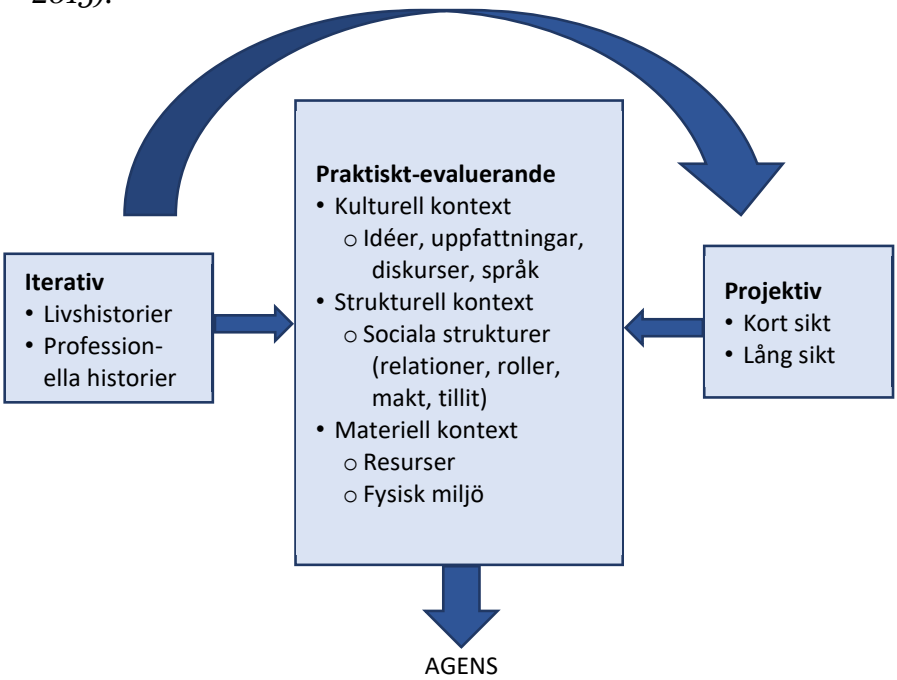
Läraragens

Agens handlar om en persons förmåga att agera i olika situationer (Priestley m.fl., 2015). Det är dock inte en inneboende kapacitet som individen har (eller inte har), utan agens är något läraren uppnår eller förvärvar (eller inte) i en viss situation. Priestley, Biesta och Robinson uttrycker det som ett ekologiskt perspektiv på läraragens. Det innebär att agens är ett resultat av samspelet mellan individers kapacitet och rådande kontextuella villkor, vilka inramar lärarens handlingsutrymme.

Teorin om läraragens (se figur 5) utvecklar ramfaktorteorin och beslutsmodellen genom att lyfta in en historisk (iterativ) och en framtidsorienterad (projektiv) dimension, tillsammans med en nutida (praktiskt-evaluerande) (Priestley m.fl., 2015). Agens uppnås inom ständigt föränderliga kontexter och orienterar mot både det förflutna, nuet och framtiden, vilka varierar mellan varje uppnådd agens.

I den iterativa dimensionen påverkas lärares agens av både den generella historieutvecklingen och den mer specifika individuella historien (Priestley m.fl., 2015). De interna faktorerna i beslutsmodellen i form av lärares uppfattningar och preferenser finns med som en del av den iterativa dimensionen, då dessa formas av lärarens individuella historia. I den iterativa dimensionen finns också den allmänna historien med i form av skolsystemets utveckling och reformer, vilket inkluderar pedagogiska ramfaktorer liksom en extern faktor (styrdokument). I både ramfaktorteori och läraragens lyfts vikten av lärares repertoarer fram i form av

Figur 5. *Ett ekologiskt perspektiv på läraragens (Priestley m.fl. 2015).*



både metoder och ämneskunskaper, vilka påverkar lärares frirum och agens (Lindblad m.fl., 1999; Priestley m.fl., 2015). Läraragens berikas av att en lärare har en bred repertoar av alternativ att välja mellan. Repertoaren påverkas i sin tur av lärarens personliga kapacitet (förmågor och kunskaper), uppfattningar (vilka kan vara både professionella och personliga) samt värderingar. Gemensamt för dessa är att de har sin grund i tidigare erfarenheter, varav lärarutbildning är en. Erfarenhet från det vardagliga lärararbetet har dock troligen störst betydelse genom inverkan från skolkultur och professionella engagemang i skolan. En svensk studie indikerar att lärares tolkning och implementering av nya reformer påverkas av erfarenheter från tidigare reformer (Ryder m.fl., 2018). En viktig skillnad mellan ramfaktorteori och läraragens är att den förra betonar ramarna inom vilka lärare kan ha ett visst frirum, medan den senare betonar lärarens agens, vilken påverkas av olika kontextuella villkor.

Den projektiva dimensionen, vilken inte alls finns med i de andra teorierna, handlar om lärares ambitioner med sitt arbete på både kort och lång sikt (Priestley m.fl., 2015). Ambitionerna kan vara positivt orienterade med intentionen att utveckla elevers lärande och välbefinnande med en önskan att sätta elevernas intressen i centrum. Ambitionerna kan också vara mer instrumentella, till exempel att upprätthålla ett

”normalt och acceptabelt klassrum”. En lärares agens kan ha stöd i styrdokument, men också stå i motsats till dem. Ambitionerna har ofta sin grund i lärares värderingar och uppfattningar. Lärarens ämnesidentitet kan också påverka.

Den praktisk-evaluerande dimensionen relaterar till lärares dagliga arbetsmiljö, vilket sker i en viss kontext (Priestley m.fl., 2015). Teorin om läraragens delar in kontexten i kulturella, strukturella och materiella faktorer. Kulturella faktorer handlar om hur idéer, värderingar, uppfattningar och diskurser tänks och uttrycks (både en inre och en yttre dialog). Materiella faktorer utgörs av resurser som möjliggör eller hindrar agens, liksom av den vidare fysiska miljön i och genom vilken agens uppnås. Strukturella faktorer relaterar till sociala strukturer i form av relationer, roller, makt och tillit, vilka påverkar agens. De kontextuella villkoren liknar ramfaktorer genom att de möjliggör viss agens (genom tillgängliga resurser) och begränsar annan (till exempel genom uppfattningar om oacceptabla risker). Detta kan innebära att lärares uppfattningar och praktik inte alltid stämmer överens. Deras agens påverkas av hur lärarna positionerar sig inom sin kontext, men också av kontextuella faktorer utom deras omedelbara kontroll. Den sociala strukturen i form av lärares relationer med andra lärare eller andra aktörer påverkar agens. Sociala strukturer handlar om hur personer positionerar sig gentemot varandra där makt och tillit är viktiga relationella resurser. De sociala strukturerna kan både möjliggöra och begränsa agens. Till exempel kan en lärare med bred erfarenhet och en stark strävan med sitt arbete, vilket ger en stark iterativ och projektiv orientering för lärarens agens, ändå hindras i sammanhang där kollaborativt arbete är begränsat.

Den praktiskt-evaluerande dimensionen har vissa likheter med ramfaktorteori och externa faktorer. Läraragensens materiella villkor motsvaras i ramfaktorteorin av resursrelaterade ramar. Strukturella faktorer liknar till viss del administrativa ramar i form av skolans ledning och arbetslag, men även organisationsrelaterade ramar som också handlar om sociala förhållanden. Den sistnämnda inkluderar också skolkultur, vilken finns med i kulturella faktorer i läraragens. Kulturella faktorer är dock ett vidare begrepp inom läraragens, där även vidare diskurser kan ingå, till exempel performativa eller formativa diskurser, vilka helt saknas i ramfaktorteorin och beslutsmodellen. En faktor som inte tydligt finns med i läraragens är de elevrelaterade ramfaktorerna med elevernas bakgrund och förutsättningar. De skulle kunna ingå i de strukturella faktorerna, men tycks inte beaktats explicit i det ekologiska perspektivet på läraragens (Priestley m.fl., 2015). Delstudierna i denna avhandling visar att detta är något lärare tar hänsyn till i sin bedömningspraktik och läraragens skulle främjas av att också beakta detta perspektiv, antingen som en del i strukturella villkor, eller som en egen kategori. Slutsatsen är att

ett ekologiskt perspektiv på läraragens mestadels innefattar perspektiven i både ramfaktorteorin och beslutsmodellen, samtidigt som den också utvidgar med fler perspektiv. Teorin om läraragens erbjuder således en bredare analysmodell som också starkare betonar lärarnas agens och därför används främst denna teori i avhandlingens diskussion, men det relateras också en del till de andra teoretiska perspektiven.

Sammanfattningsvis uppnår lärare agens när de har förmåga att välja mellan olika alternativ i en viss situation och kan bedöma vilket alternativ som är mest önskvärt i relation till de syften som gäller för den praktik i vilken de agerar (Priestley m.fl., 2015). Om alternativen är få begränsas agensen. Agens saknas om det inte finns några handlingsalternativ eller om lärare bara följer rutiner eller vanor utan att överväga olika alternativ. Detta gäller generellt i lärares arbete liksom i bedömningsprocessen.

4 Tidigare forskning

Denna avhandling handlar om lärares bedömningspraktik i samhällskunskap på gymnasiet. Med bedömningspraktik avses lärares syften med sina bedömningar, deras beslut om vilket innehåll och vilka förmågor som ska bedömas samt med vilka metoder detta ska mätas. Enligt en forskningsöversikt har intresset för forskning om lärares arbete med bedömningar varit svagt under 1900-talet i Sverige (Forsberg & Lindberg, 2010; Lundahl m.fl., 2015). Sedan 2000 har intresset ökat och en hel del forskning har tillkommit men lärares bedömningspraktik är fortfarande ett relativt outforskat område. Det finns lite mer internationell forskning om lärares bedömningspraktik men en internationell kunskapsöversikt pekar ut ett fortsatt behov av mer forskning inom området (McMillan, 2013). Denna avhandling bidrar med kunskap till detta fält.

Den forskning som redovisas nedan är den som berör lärares bedömningspraktik (gällande syften, innehåll och metoder) i klassrummet, vilket är studiens fokus. Översikten berör forskning om betygsättning, då ett viktigt och vanligt syfte är att använda bedömningar som underlag för betyg, men dock bara kort då studiens fokus är mer på processen i vad som sker i klassrummet. Det finns gott om forskning om formativ bedömning (Lindberg & Hirsh, 2015), särskilt om återkoppling, kamrat- och självbedömning, medan intresset för vilka uppgifter lärare använder som underlag för formativ bedömning inte har varit lika stort. Forskning om lärares uppgifter som används för formativ bedömning tas upp nedan, då det är av intresse för avhandlingens syfte. Nationella prov förekommer inte i ämnet samhällskunskap på gymnasiet och därför tas inte heller forskning om externa prov upp.

Först redogörs för olika typer av bedömningar och forskning om detta. Därefter redovisas forskning om lärares bedömningspraktiker i klassrummet, följt av forskning om bedömning i olika ämnen respektive specifikt i samhällskunskap. En av forskningsfrågorna handlar om faktorer som har betydelse för lärares bedömningspraktik och forskning om detta redovisas i ett avsnitt. Därefter redovisas forskning om konsekvenser av lärares bedömningspraktik, inklusive forskning om rättvisa och likvärdighet, som också är av relevans för denna avhandling.

Olika typer av bedömningar

Två viktiga syften med bedömning är summativ respektive formativ bedömning, vilka också har betydelse för lärares beslut om bedömningar (Dolin m.fl., 2018). Därför redovisas i detta avsnitt forskning om dessa olika typer av bedömningar Fokus är det som relaterar till lärares

bedömningspraktik i form av de uppgifter de använder (som de ofta konstruerar själva) för att bedöma elevers kunskaper, oavsett om det är i summativt eller formativt syfte. Det innebär att forskning om formativ bedömning gällande återkoppling, kamrat- och självbedömning inte tas upp.

En viktig aspekt att vara uppmärksam på är att formativ och summativ bedömning handlar om olika *syften* med bedömning och inte olika *former* av bedömning (Dolin m.fl., 2018). Samma bedömningsuppgifter kan oftast användas som underlag i både formativa och summativa syften. I en amerikansk enkätstudie med 140 lärare från årskurs 3 till och med gymnasiet, var det dock ovanligt att bedömningsuppgifter används för enbart formativa syften (endast ca 12 procent), medan omvänt de flesta bedömningar gjordes i slutet av ett arbetsområde (ca 75 procent) och användes som underlag för betyg (ca 88 procent) (Frey & Schmitt, 2010). Samtidigt finns det en hel del studier som pekar på problem med att blanda summativ och formativ bedömning (Lauvås m.fl., 2019; Shute, 2008). Det finns också ganska mycket forskning som relaterar till summativa och formativa syften och diskuterar distinktionen mellan dem (Lindberg & Hirsh, 2015) och en del undersöker också lärares användning av dessa (se till exempel Leong, 2014). Det är mindre vanligt att också undersöka andra syften. Det finns ett antal studier som handlar om lärares uppfattningar om bedömningar (se till exempel Brown, 2004; Harris & Brown, 2009) men dessa handlar om syften med bedömningar på en systemnivå och inte lärares syften med sina bedömningar på klassrumsnivå.

I summativ bedömning är syftet att bedöma elevens kunskaper vid en viss tidpunkt, främst i slutet av ett arbetsområde, och summativ bedömning definieras ofta som bedömning *av* lärande (Dolin m.fl., 2018; Wilson, 2018). Den summativa bedömningen används sedan för beslut om till exempel särskild undervisning eller för att informera om elevens kunskapsutveckling till vårdnadshavare. Ett vanligt syfte i summativ bedömning är att uppgifter används som underlag för betyg. I formativ bedömning är syftet att bedömningar ska bidra till elevens kunskapsutveckling, vilket uttrycks som bedömning *för* lärande (Dolin m.fl., 2018; Jönsson, 2023a). Ibland uttrycks det också som bedömning *som* lärande, vilket innebär att även själva bedömningsuppgiften ska bidra till elevens lärande. Viktiga komponenter i formativ bedömning är att läraren formulerar tydliga lärandemål (vart eleven ska), samlar in underlag om elevernas kunskaper, tolkar och utvärderar elevernas prestation samt ger återkoppling till eleverna om både var de befinner sig och hur de ska gå vidare. Utifrån resultaten anpassar också läraren sin undervisning för att främja elevernas fortsatta lärande. Ytterligare viktiga inslag i formativ bedömning är kamratbedömning och självbedömning. Av komponen-

terna i den formativa processen är det insamlandet av underlag om elevernas lärande, vilken handlar om vad lärare bedömer och med vilka metoder, som är av intresse för denna avhandling.

Forskningen om hur lärare samlar in underlag om elevernas kunskapsutveckling i formativt syfte har främst berört hur lärare arbetar med frågor i klassrummet (Heritage, 2013; Penuel & Shepard, 2016). För att kunna användas formativt betonar forskningen att frågorna behöver vara öppna så att de lockar fram elevernas tänkande och därför inte vara faktafrågor där det finns ett rätt eller fel svar. Andra metoder som forskningen tar upp är vikten av observationer i klassrummet, då det också är viktigt att läraren ställer frågor om elevernas arbete för att få insikt i elevens tänkande om uppgifterna. I litteratur om formativ bedömning lyfts autentiska, det vill säga verklighetstrogna, uppgifter fram som önskvärda (se till exempel Jönsson, 2023a).

När summativ bedömning används som underlag för betyg är det viktigt att lärares bedömningar grundas på relevanta underlag (Dolin m.fl., 2018; McMillan, 2018; Wilson, 2018). Säkerheten i summativa bedömningar påverkas av kvaliteten i bedömningsuppgifterna (Moss, 2013). Till exempel behöver bedömningsmetoder matcha målen för undervisningen för att det ska bli samstämmighet mellan kursmål, undervisning och bedömning, vilket ger lärarna möjlighet att göra valida bedömningar av elevernas kunskapsnivå (McMillan, 2018).

Samtidigt har en del forskning pekat på en inkonsekvens mellan lärares uppfattning om validitet och deras summativa praktik, med en diskrepans mellan bedömningarnas syfte och hur lärarna genomför sina bedömningar (Black m.fl., 2010). Orsaken till detta menar Black m.fl. är dels externa testers negativa inverkan på lärares bedömningspraktik, dels att validitet inte är så framträdande i lärarnas tänkande. Ytterligare en aspekt som kan förklara denna diskrepans är olika kontextuella villkor som omger lärares arbete, vilket tas upp i denna avhandling.

Mycket forskning om summativ bedömning har handlat om lärares betygsättning (Brookhart, 2013). Denna avhandling behandlar inte själva betygsättningen men eftersom många av lärarna använder sina bedömningar som underlag för betyg är det relevant att relatera till sådan forskning. Denna forskning har till stor del handlat om vilka underlag som använts vid betygsättning och den har visat att många lärare uttryckt att det är lämpligt att väga in icke-akademiska faktorer som ansträngning och hur eleven utvecklats då detta hjälper eleverna att lyckas (Bonner, 2016; Brookhart, 2013; Klapp, 2016; Selghed, 2011). Forskning har hittills inte visat att kriterierelaterad betygsättning skulle ha påverkat vilka faktorer lärare väger in i betygen men däremot att lärare upplever en spänning mellan sina uppfattningar om formativ bedömning och förväntningarna på att genomföra summativa bedömningar. I betygsätt-

ningen strävade lärarna efter att vara rättvisa (Brookhart, 2013). Vidare fanns en individuell variation i lärares betygsättningspraktik och lärare hade olika uppfattningar och tolkningar av syftet med betygen. Brookhart sammanfattade med att förståelsen av orsakerna till lärares betygsättningspraktik är komplex, då betyg har flera innebörder och sker i flera kontexter, inklusive sociala och legala kontexter.

Oavsett om lärares syfte är formativt, summativt eller både och behöver lärare utforma uppgifter för att kunna bedöma elevernas kunskapsutveckling (Dolin m.fl., 2018). Syftet med uppgifterna i formativ bedömning är att ge underlag för att utvärdera var eleven befinner sig i sin kunskapsutveckling och för att kunna ge eleven återkoppling för att utveckla lärandet vidare (Heritage, 2013). Därmed blir en viktig validitetsfråga huruvida bedömningarna faktiskt bidrar till detta (Dolin m.fl., 2018). Uppgifterna behöver då utformas med en tydlig koppling till de aktuella lärandemålen, så att läraren kan göra valida bedömningar och ge relevant återkoppling. I summativ bedömning behöver bedömningsuppgifter matcha målen/kriterierna så att läraren kan dra valida slutsatser om elevernas kunskapsnivå, och därmed ha goda grunder för att sätta betyg. Denna studie handlar inte specifikt om formativ eller summativ bedömning men hanterar dessa syften i relation till lärares bedömningspraktik när det gäller val av innehåll och förmågor som bedöms samt med vilka metoder. Dessa relationer går i båda riktningarna, det vill säga såväl formativa som summativa syften påverkar lärares bedömningspraktik samtidigt som bedömningspraktiken har betydelse för den formativa och summativa bedömningen.

Lärares bedömningspraktiker

I detta avsnitt redovisas forskning om hur lärares bedömningspraktiker ser ut och hur lärare resonerar om sina beslut om bedömningar när det gäller val av innehåll och förmågor som ska bedömas samt med vilka metoder.

Det finns en hel del litteratur om hur bedömningar bör göras (se till exempel Chappuis & Stiggins, 2019; McMillan, 2018) men enligt Torrez och Claunch-Lebsack (2014) är det brist på forskning om lärares bedömningspraktik. Det finns således information om hur lärare bör göra, men inte så mycket om hur de faktiskt gör, trots att bedömning är en så central del i lärares arbete. I USA gjordes en del forskning om detta under 1980-talet fram till tidigt 2000-tal men därefter har intresset för att studera lärares bedömningsuppgifter varit svalt (Wellberg, 2023).

När det gäller kunskapsformer har forskningen visat att det främst är minneskunskaper som testas (Frey & Schmitt, 2010; Mattson, 1989; Moss, 2013). Högstadielärare uppgav dock att de både använder frågor

som testar minneskunskaper och uppgifter som testar resonemang och förståelse (Brookhart, 2013; Mertler, 1999; Zhang & Burry-Stock, 2003) men det var minneskunskaper som dominerade. Detta gällde generellt men med en viss variation mellan ämnen, vilket utvecklas i ett särskilt avsnitt nedan.

Dessa minneskunskaper har främst testats med skriftliga prov (Mattson, 1989; Moss, 2013). En amerikansk enkätstudie med 140 lärare visade att skriftliga prov var den vanligaste bedömningsformen, främst med flervals-, kortsvars- och essäfrågor (Frey & Schmitt, 2010), även om det enligt studien hade blivit något vanligare med tillämpningsuppgifter som grupparbeten, presentationer samt essä- och inlämningsuppgifter sedan 1980-talet. I en annan enkät- och intervjustudie uppgav 155 amerikanska lärare också att de använde traditionella papper-och-penna-prov trots att de föredrog att använda andra alternativa metoder som inlämningsuppgifter, rapporter och grupparbeten (Bonner m.fl., 2018).

Olika internationella studier har visat att det fanns olikheter i lärares bedömningsmetoder mellan olika stadier. Den generella bilden i USA, baserat på olika lärarenkätstudier, var att det i låg- och mellanstadiet användes mer informell bedömning, inklusive observationer, medan det i högstadiet och gymnasiet var vanligare med skriftliga uppgifter, främst egenproducerade prov (Brookhart, 2013; Mertler, 1999; Zhang & Burry-Stock, 2003). Men det förekom även andra uppgifter som laborationer, klassrumsarbete och läxor (Wherfel m.fl., 2022; Zhang & Burry-Stock, 2003). Lite mer sällan användes grupparbeten och uppsatser. Det kan dock skilja sig lite mellan länder. I en enkätstudie från Ghana uppgav lärare i lägre stadier att de i högre grad använde prov med objektiva frågor jämfört med högstadielärare, vilka i högre grad använde essäuppgifter (Osman, 2021). Till skillnad från lärarna i de amerikanska studierna visade en nederländsk studie med tolv högstadie- och gymnasielärare en större variation av uppgifter, som papper-och-penna-prov, muntliga frågor och prov, inlämningsuppgifter, observationer och presentationer (Kippers m.fl., 2018). Dessa olika resultat mellan länderna tyder på att lärare utvecklar olika bedömningspraktiker i olika nationella kontexter, vilket utvecklas mer nedan i avsnittet om kontextuella villkor.

I en amerikansk studie med 24 grundskolelärare hade lärarna svårt att förklara grunderna för sin bedömningspraktik, särskilt om de hade arbetat länge (McMillan & Nash, 2000). Lärarnas bedömningspraktiker var starkt personligt präglade och de såg bedömning som något privat som sällan diskuterades med kollegor. Lärarna föredrog sina egna konstruerade bedömningsuppgifter. En aspekt som påverkade deras val av metoder var vilken typ av kunskap som skulle bedömas. Om det till exempel var minneskunskaper så användes frågor med fasta svarsalternativ, medan om det var mer förståelse, tillämpning eller resonemang så

användes öppna frågor. Lärarna menade att det bästa var att använda en variation av uppgifter för att bedöma lärande. Lärarna uttryckte en stark önskan att eleverna skulle lyckas och utvecklade praktiker för att hjälpa dem.

Även 126 finska lärare uppgav i en enkätstudie att det är positivt att använda olika bedömningsmetoder (Atjonen, 2014). Lärarna lyfte också fram vikten av tydliga kriterier och tydliga underlag för bedömningen och att diskutera med eleverna och ge återkoppling då det bidrar positivt till bedömningen. Däremot uttryckte de att olämpliga bedömningsmetoder som användning av poäng är negativt då det inte bidrar till lärande. De uttryckte också att bedömningar kan misslyckas om uppgifterna är bristfälligt konstruerade. Enligt Atjonen kan bristfälliga kunskaper vara en orsak till svaga bedömningsunderlag.

Sammanfattningsvis är bilden att lärares bedömningspraktik har präglats av skriftliga prov som testar minneskunskaper, även om senare forskning tyder på en viss förändring mot en ökad variation. Lärares bedömningspraktiker skilde sig också mellan olika stadier i skolan, där den generella bilden är att det var vanligare med faktaprov högre upp i skolan. Detta varierade dock något mellan länder och bedömning tycks skilja sig mellan olika nationella kontexter. Samtidigt tycks bedömningsarbetet vara individuellt och lärare hade svårt att förklara varför de gör som de gör. Det är således intressant att studera lärares bedömningspraktiker i olika länder och på olika stadier samt deras rationaler för dessa praktiker. Många av de studier som gjorts bygger på lärares självbedömningar i enkäter eller intervjuer. Denna avhandling bidrar med kunskap till detta fält genom att studera bedömningspraktiker på gymnasiet i Sverige. Delstudierna använder också en blandad metod med både intervjuer och analyser av lärares egna bedömningsuppgifter, vilket ger ett rikare underlag.

Bedömning i olika ämnen

Denna avhandling handlar om bedömning i ett specifikt ämne. Olika skolämnen har olika karaktär, vilket forskning har visat kan påverka hur bedömningar görs i olika ämnen. Det finns dock inte så många studier om hur ämnets karaktär påverkar bedömningar (McMillan, 2013). Tidigare studier indikerar emellertid att ämnets karaktär påverkar lärares bedömnings- och betygsättningspraktik (Biberman-Shalev m.fl., 2011; McMillan, 2001; Prøitz, 2013) och att det finns en skillnad i bedömningspraktik mellan olika ämnesområden (Zhang & Burry-Stock, 2003).

Ämnens karaktär kan vara antingen öppna och flexibla med en svag inramning eller slutna och hierarkiska med en stark inramning (Bernstein, 1975; Biberman-Shalev m.fl., 2011). I en israelisk enkätstudie på

gymnasiet visades att i ämnena hebreiska/arabiska och naturvetenskap, vilka uppfattades som mer öppna och flexibla, tenderade lärarna att väga in elevernas ansträngning i bedömningen i större utsträckning än i matematik som uppfattades som slutet och hierarkiskt, där elevernas prestation var mer avgörande. Biberman-Shalev m.fl. menar dock att ett ämne kan uppfattas olika i olika länder och i Sverige beskrivs naturvetenskapliga ämnen som slutna/hierarkiska med stark inramning, medan samhällskunskap betraktas som ett öppet och flexibelt ämne med svag inramning (Linde, 2021). En amerikansk enkätstudie på högstadiet fann också att lärare i samhällskunskap tenderade att väga in ansträngning, närvaro och läxor i större utsträckning än lärare i engelska och matematik (McMillan, 2001), vilket är i linje med mönstren i den israeliska studien.

En amerikansk enkätstudie med 1483 högstadie- och gymnasielärare i naturvetenskap, engelska, matematik och SO (social studies) visade en tendens till att lärares bedömningspraktik varierar mellan ämnen (McMillan, 2001). De som främst skilde ut sig var engelsklärare, vilka tenderade att i sina bedömningsuppgifter betona förståelse, tillämpning och resonerande mer än lärarna i andra ämnen. De använde i högre grad uppgifter med öppna frågor, medan de andra mer använde minnesfrågor och bundna frågor (flervalsfrågor). Skillnaderna var dock små och ytterligare studier efterfrågades. I en annan enkätstudie framkom att lärare i språk i större utsträckning använde tillämpningsuppgifter än lärare i matematik, naturvetenskap och SO (social studies) (Frey & Schmitt, 2010). Även i Sverige tycks språklärare använda en variation av bedömningsuppgifter. I en studie, baserad på både enkäter och intervjuer, uppger lärare i engelska och moderna språk, både på högstadiet och gymnasiet, att de till stor del grundar sin betygsättning på observationer och fri muntlig kommunikation i klassrummet, tillsammans med lärarkonstruerade tester och uppsatser (Oscarson & Apelgren, 2011).

I en svensk studie med 25 gymnasielärare i naturvetenskapliga ämnen dominerade prov som bedömningsform och de användes främst som underlag för betyg (Gomez, 2017). Lärarna använde även andra uppgifter som läxor och labbrapporter, men samtliga lärare uppgav att prov var det främsta betygsunderlaget. Lärarna tänkte sig att proven inte bara testar minneskunskaper utan även andra kognitiva förmågor som procedurala och analytiska förmågor liksom kritiskt tänkande. De andra uppgifterna användes främst som komplement vid osäkra fall. Muntliga frågor var främst kontrollerande faktafrågor och eleverna uppmuntrades inte till att ställa egna frågor (snarare var tystnad ett ideal) eller att diskutera naturvenskap. För lärarna var bedömning och betygsättning synonymt. Deras fokus var således på summativ bedömning och formativa strategier som kamratbedömning eller självbedömning användes inte. Prov i

teknik på högstadiet visade sig i en annan svensk studie (av 30 prov från 12 olika grundskolor) också vara faktaorienterade med kortsvarsfrågor och begränsade essäfrågor och där resultaten ofta summerades med poäng (Hartell & Strimel, 2019). Underlaget var dock relativt litet och författarna poängterade att lärarna också använde andra bedömningsuppgifter, vilka inte undersöktes. I matematik dominerade också summativ bedömning med skriftliga prov som främsta underlag i en intervjustudie med 15 lärare (Becevic, 2011). Även matematiklärarna använde en del andra uppgifter men liksom lärarna i naturvetenskapliga ämnen användes de som komplement till proven för att avgöra betygen i gränsfall.

Andra studier har visat en liknande bedömningskultur med fokus på faktafrågor i biologi på högstadiet i Turkiet (24 lärare) (Cimer & Cimer, 2010) och Bangladesh (30 lärare) (Rahman, 2018). De turkiska lärarna använde främst faktainriktade prov, medan lärarna i Bangladesh främst använde muntliga och slutna faktafrågor, vilka enligt båda studierna främjade minneskunskaper. (Individuella skriftliga uppgifter var också vanligt men även dessa testade främst ytligare faktakunskaper.) De turkiska lärarna uttryckte dock att de testar andra förmågor men i praktiken testades främst minneskunskaper. Det fanns således en diskrepans mellan vad de sade att de gör och deras praktik. En intressant notering är att lärarna i den turkiska studien uttryckte att de undvek frågor som testar mer resonerande förmågor utifrån att lärarna hade en uppfattning om att eleverna inte klarar sådana uppgifter. I stället för att främja att eleverna utvecklar sådana förmågor, förstärkte lärarna elevernas lärstrategier mot minneskunskaper.

Två kvalitativa svenska studier visade att historielärarna på gymnasiet hade en ganska låg överensstämmelse (*alignment*) mellan kursplanen i historia och sina bedömningsuppgifter, avseende både innehåll och förmågor som testades (Kakoulidou, 2021; Rosenlund, 2016). Kursplanen i historia förändrade fokus 1994 från att vara faktaorienterad till att bli mer empiriskt orienterad och med betoning på historiska kompetenser. Detta tycks dock inte ha slagit igenom i de studerade lärarnas bedömningspraktiker, även om en viss förändring tycks ha skett mellan studierna. Studierna visade att historielärarna tenderade att testa mer av att minnas och förstå faktakunskaper än resonerande förmågor. Vissa förmågor i historia testades inte alls av en del lärare. Lärarna i Rosenlunds studie uttryckte att deras undervisning och bedömningspraktik formats av en tradition som förmedlats genom deras egen skolgång och från mer erfarna kollegor, vilket är i linje med annan forskning (Cooper m.fl., 2017; Edwards, 2020; Forsberg & Wermke, 2012; Looney m.fl., 2018). Ett ämnes praktiker kan konserveras genom överföringen mellan

olika lärargenerationer (Elwood & Murphy, 2015) och det tycks finnas en stark sådan tradition i ämnet historia.

Ämnets karaktär och inramning tycks således ha betydelse för hur bedömningspraktiker formas i olika ämnen. Det är därför av intresse att studera bedömningspraktiker i olika ämnen. I många ämnen tycks det vara faktakunskaper som betonas, vilka ofta testas i traditionella skriftliga prov. I denna avhandling studeras bedömning i samhällskunskap och forskning om detta redovisas i nästa avsnitt.

Bedömning i samhällskunskap

I det svenska samhällskunskapsämnets ämnes- och kursplaner framträder medborgarfostran, samhällsvetenskaplig metod och reflekterande frågor som en tydlig karaktär i ämnet (Möllenborg, 2023; Wahlström m.fl., 2020). Samtidigt menar Wahlström m.fl. (2020) att flera av läroplanens generella förmågor också är framskrivna i ämnet och att dessa generella förmågor är mer betonade än ämnesinnehållet. De menar att dessa generella förmågor utgör en stark inramning i ämnet på grund av ett starkt fokus på kunskapskraven och därmed kraven på att eleverna måste visa upp sina kunskaper för läraren. Möllenborgs (2023) studie visade dock en annan bild då lärarna i den studien uttryckte att de använder analyser i relation till samhällsfrågor och ofta med specifika samhällsvetenskapliga analysmetoder. Även om ”analysera” är framskrivet som en generell förmåga i läroplanen tolkas och används den ämnesspecifikt av lärare i den faktiska praktiken.

Det finns inte så mycket forskning om bedömning i samhällskunskap, varken internationellt (Barton & Avery, 2016; Torrez & Claunch-Lebsack, 2014) eller i Sverige (Berg & Persson, 2020). Däremot finns det en del litteratur som talar om hur bedömningar borde utformas i ämnet. I både amerikansk och svensk samhällskunskapsdidaktisk litteratur betonas att ämnets analytiska och problemlösande karaktär innebär att lärare bör undvika att använda slutna och faktabaserade frågor och i stället använda mer öppna frågor som stimulerar elevernas eget tänkande (Arevik & Hartzell, 2015; Barton & Avery, 2016; Lindqvist & Hyltegren, 1995; Svingby, 1989; Torrez & Claunch-Lebsack, 2014). Alternativa bedömningar uppmuntras i stället för traditionella papper-och-penna-prov. Bedömningar bör också vara autentiska och inriktade mot tillämpning.

Lärares bedömningspraktik tycks dock skilja sig från litteraturens ideal. I en enkätstudie (Brookhart & Durkin, 2003) bland gymnasieelever visades att elevernas bild utifrån lärarnas bedömningsuppgifter är att de ska lära sig innehållet i läroboken och lärarens undervisning. Samhällskunskapslärare skilde sig inte från andra ämnen utan även de testade främst faktakunskaper i form av frågor med fasta svarsalternativ

i stället för att använda tillämpningsuppgifter (*performance assessments*) (Curry & Smith, 2017; Mattson, 1989).

Det finns få svenska studier om bedömning i samhällskunskap, i både grundskolan och gymnasiet. En av få studier är en omfattade enkätstudie som gjordes under det normrelaterade betygssystemet (Mattson, 1989). Studien bygger på ca 2400 enkäter där lärare i olika ämnen fick kategorisera sina skriftliga prov. I de flesta ämnen dominerade faktafrågor, så även i SO-ämnen (95% på högstadiet och 100% på gymnasiet). Enligt Måhl (1991) skulle detta kunna bero på att de normrelaterade betygen som syftade till rangordning i sig orsakade problem med reproducerande faktafrågor. Problemet tycks dock inte helt ha försvunnit med de mål- och kunskapsrelaterade betygen som infördes 1994. I en studie av lokala kursplaner visar Tholin (2003) att dessa innehållsligt ofta var traditionella och fokuserade på det som är mätbart som faktakunskaper. Fyra samhällskunskapslärare uttryckte dock i en intervjustudie (Bernmark-Ottosson, 2009) att de i början använde mer av lätträttade faktaprov men att de med tiden använde mer öppna och analyserande frågor. Lindmark (2013) undersökte samhällskunskapslärares konceptioner om ämnet och hur dessa bedömdes. Han fann i en enkätundersökning med 60 lärare fyra konceptioner: den fakta- och begreppsinriktade, den värdegrundsinriktade, den samhällsanalytiskt inriktade och den medborgarfärdighetsinriktade ämneskonceptionen. En uppföljande intervjustudie med åtta lärare gjordes i kombination med analyser av några av deras skriftliga bedömningsuppgifter visade att lärare ofta betonade medborgar- och värdefokus men att detta sällan bedömdes. I stället dominerade fakta- och begreppsfrågor.

Skolverket (Svingby, 1998) samlade in cirka 400 prov i en nationell utvärdering av grundskolan[6], inklusive SO-ämnen. Resultatet visade att 95 procent av uppgifterna i proven var reproducerande faktafrågor, vilka premierade minneskunskaper och ”lokaliseringsläsning” (att leta efter något bestämt) i läroboken. Elevernas reflektion, kreativitet och egna tankar efterfrågades inte i proven och slutsatsen var att proven inte var i linje med läroplanens mål. I en enkätstudie med 127 lärare (Törnvall, 2001) uttryckte lärare i olika ämnen att de främst använder prov i summativt syfte. Därför föredrog de lätträttade faktaprov där resultaten är lätta att presentera i siffror, medan de undvek essäfrågor och problemlösande frågor som kräver subjektiva bedömningar. Samtidigt uttryckte lärarna i studien en konflikt utifrån att läroplanen kräver en annan bedömningspraktik.

[6] Detta är inte en forskningsstudie, utan en Skolverksrapport. Resultaten är dock av intresse och därför tas den upp i denna översikt.

Parallellt med denna avhandlings delstudie I undersökte Odenstad (2011) bedömning i samhällskunskap genom att samla in och analysera 95 prov från 23 samhällskunskapslärare. Lärarna fick också besvara en enkät: Där uttryckte de bland annat en summativ bedömningspraktik där prov utgör den främsta grunden för betygen. Hennes studie visade en skillnad i lärares bedömningspraktik mellan studieförberedande och yrkesförberedande program. På yrkesförberedande program tenderade proven att väga något mindre tungt vid betygsättningen och den muntliga aktiviteten på lektionerna hade lite större betydelse, särskilt för de högre betygen. Proven tenderade att vara mindre omfattande i både omfång och kunskapsinnehåll. Även frågornas karaktär skilde sig. I båda typerna av program dominerade faktafrågor men i än högre grad på de yrkesförberedande programmen där det även var vanligare med frågor som kräver kortare svar. Vidare var proven ofta nivågrupperade med mer faktafrågor för betyget Godkänt och förståelsefrågor för de högre betygen (VG och MVG). De insamlade proven handlade främst om Sveriges styrelseskick och samhällsekonomi medan några prov handlade om EU, internationell politik, massmedia och kriminologi. Faktafrågor dominerade generellt men mer om hur Sverige styrs. Det förekom fler förståelsefrågor om de andra ämnesområdena. Det tycks således finnas en kärna i vilka ämnesområden som betonas i lärares bedömningsuppgifter (Odenstad, 2011; Öberg & Bäckström, 2021). Det är främst statsvetenskap och samhällsekonomi som har betonats i lärares bedömningar.

Sammanfattningsvis har bedömning i samhällskunskap dominerats av prov som testar faktakunskaper, trots att ämnets karaktär är en orientering mot mer resonerande förmågor och kritiskt tänkande. Det har dock gått en tid sedan dessa få studier genomfördes och det har sedan dess genomförts en del skolreformer med bland annat nya betygskriterier. Det är därför av intresse med fler och nyare studier om bedömning i samhällskunskap med dess öppna och flexibla karaktär.

Bedömningspraktikers kontextuella villkor

En av avhandlingens frågor handlar om faktorer som kan påverka lärares arbete med bedömningar. Lärares bedömningspraktik formas i ett sammanhang där olika kontextuella villkor har betydelse och i detta avsnitt redovisas forskning härom.

En sådan kontext är ett lands kultur. Forskning har visat att bedömningspraktiker skiljer sig mellan lärare i olika länder (Forsberg & Wermke, 2012; Klenowski, 2013). Det är också rimligt mot bakgrund av att bedömningar är situerade och utformas i en historiskt framväxt diskurs (Allal, 2013; Elwood & Murphy, 2015). Kulturella uppfattningar om bedömningar påverkar, till exempel vad bedömningar är och vad de är till

för, synsätt på lärande, men också på elever. Dessa uppfattningar överförs från generation till generation och tenderar därmed att konserveras i bedömningars syften, användning och praktiker.

En annan faktor är lärares identitet som bedömare, vilken har betydelse för lärares bedömningspraktik (Looney m.fl., 2018). Denna identitet konstrueras under tiden som student och som yrkesverksam genom interaktion med andra och genom rollen som bedömare. Bedömaridentiteten formas av lärares uppfattningar, känslor, kunskaper och färdigheter i relation till bedömning. Lärares uppfattningar påverkas av externa (till exempel läroplan, systemkrav på bedömning och skolans organisation) och interna (lärarens egna uppfattningar, värderingar, bakgrund och erfarenheter) faktorer, vilket påminner om McMillans (2003) modell om beslutsfattande. Nedan tas några av dessa aspekter upp: utbildning, erfarenheter, kollegor, uppfattningar och känslor.

Allt detta formar också lärares bedömarkompetens. Det finns olika definitioner på vad som ingår i lärares bedömarkompetens eller bedömarliteracitet som det ofta också benämns (Pastore & Andrade, 2019). I en del länder, till exempel USA, Australien och Storbritannien, finns det nationella standarder för vad som ingår i lärares bedömarkompetens. Ofta är fokus på vad lärare behöver kunna göra i praktiken, som att välja relevanta bedömningsmetoder i relation till bedömningens syfte. I nyare forskning om detta har även sociokulturella och sociopolitiska aspekter lyfts fram som viktiga kompetenser. Bedömningsliteracitet ses som en situerad praktik med ett komplext samspel mellan olika komponenter relaterade till sociala, kulturella, policy-, professionella och erfarenhetsbaserade faktorer. Utifrån detta har Pastore och Andrade utvecklat en tredimensionell modell för bedömningsliteracitet med en konceptuell, en praxeologisk samt en socioemotionell dimension. Den konceptuella dimensionen handlar om en teoretisk kompetens om till exempel begrepp och modeller. Den praxeologiska dimensionen är praktisk kompetens i att utforma och genomföra bedömningar. Den socioemotionella dimensionen innefattar en social kompetens som att kommunicera med olika intressenter men också medvetenhet om bedömarrollen, etiska aspekter och bedömningars påverkan och konsekvenser.

Utbildning skulle kunna bidra till att utveckla lärares bedömarkompetens men flera internationella studier visar att lärare ofta inte får någon eller mycket lite utbildning om bedömning och inte heller får utbildning om detta i sin yrkespraktik (Amani m.fl., 2021; Campbell, 2013; Forsberg & Wermke, 2012; Moss, 2013). Forskning har visat att lärare ofta använder praktiker som inte är i linje med rekommendationer om bedömning och det finns en diskrepans mellan lärares uppfattning om bedömning och deras praktik (Moss, 2013). Det finns också en diskrepans mellan lärares upplevda och faktiska bedömningskompetens.

Många lärare väger in irrelevanta faktorer som beteende, attityd, ansträngning eller motivation. Samtidigt, trots brister i lärarutbildningen, använder ändå lärare regelbundet en variation av bedömningssätt.

Även svenska lärare svarade i en enkätstudie att de inte har fått så mycket utbildning om bedömning under sin lärarutbildning (Forsberg & Wermke, 2012). De uttryckte att de främst hade utvecklat sin kompetens genom den egna erfarenheten, men också genom kollegor. De uttryckte även att självstudier i någon grad hade bidragit. Denna studie genomfördes dock före de svenska skolreformerna 2011 och det var först då som kurser i bedömning blev obligatoriska inom lärarutbildningen (Prop. 2009/10:89). Många lärare tycks således vara oförberedda och ha otillräckliga kunskaper om bedömning, vilket riskerar att leda till att summativa bedömningar blir oprecisa och med låg reliabilitet (Amani m.fl., 2021; Moss, 2013). Samtidigt efterlyser många lärare mer utbildning i bedömning (Moss, 2013) och forskning har visat att kurser om bedömning kan ge positiva effekter på lärares förmågor att genomföra bedömningar i linje med rekommendationer (Koloi-Keaikitse, 2016; Zhang & Burry-Stock, 2003).

Många lärare utrycker dock att de har god kompetens i bedömning men att de har lärt sig genom erfarenhet (trial and error) som lärare men också genom erfarenheter från bedömningar under universitetsstudierna (Moss, 2013). Erfarenhetens betydelse för lärares bedömningspraktik bekräftas i en mindre amerikansk enkätstudie med 87 lärare (Unal & Unal, 2019). Den visade att erfarna lärare tenderade att värdera bedömningar mer än lärare med mindre erfarenhet. Vidare tenderade mer erfarna lärare att i något större utsträckning använda uppgifter som de själva konstruerade, medan mindre erfarna oftare använde färdiga uppgifter från läromedelsförlag och liknande.

En kanadensisk fokusgruppsstudie med 76 grundskollärare visade att lärare snarare tar del av kunskaper om bedömning via kollegor än från forskning (Cooper m.fl., 2017). Brist på tid uppgavs som skäl, men även att det är svårt att hitta forskning och att förstå den. Lärarna uttryckte också att om forskningen överensstämmer med deras uppfattningar om lärande och bedömning så är det lättare att ta till sig, men om den står i strid med egna uppfattningar är det risk att forskningen avfärdas. Lärarna uttryckte att forskningsöversikter, videor, engagerande hemsidor och möjlighet att dela idéer med kollegor skulle öka möjligheterna att ta del av forskning.

Bedömningspraktiker påverkas även av den verksamhetsförlagda utbildningen (VFU). De påverkas främst av lärarstudenternas egna uppfattningar och värderingar om bedömning enligt en burmesisk interventionsstudie med 30 lärarstudenter (Oo m.fl., 2021). Samtidigt kan deras agens gällande bedömning begränsas av deras handledare beroende på

hur mycket kontroll handledaren utövar eller om de ger studenten stor frihet att utforma sina bedömningar.

Dessa olika studier bekräftas av en studie om vad som formar lärares bedömningspraktik under deras första år som lärare (Edwards, 2020). Lärarnas bakgrund och tidigare egna erfarenheter av bedömningar visade sig även i denna studie ha stor betydelse, tillsammans med intryck från kontexten, främst kollegor, under de första åren som lärare. Erfarenheterna kunde även vara känslomässiga, till exempel kunde lärare undvika uppgifter som de själva känt nervositet inför som elever.

Lärares bedömningsbeslut tycks inte bara grundas på rationella grunder, utan också på intuition enligt en studie (Vanlommel m.fl., 2017). Intuition vid bedömning relaterade till tre aspekter: a) igenkännande, b) känsla, samt c) bias. Utifrån lärares erfarenheter kände lärarna igen olika mönster och tolkade olika signaler utifrån dessa. De fattade också beslut utifrån magkänsla. Intuitiva beslut har ofta också en bias, framför allt utifrån att de reproducerar och förstärker de redan existerande uppfattningarna. Lärare tenderade att leta efter information som bekräftade det som de redan trodde och undvek information som motsade detta, eller försökte ge den andra förklaringar.

En annan faktor som kan påverka bedömningspraktiker är rådande diskurser. Priestley m.fl. (2015) menar att performativitet, kravet på att skolor och lärare ska prestera, påverkar lärares agens. Lärare måste hantera dilemman mellan hur de ska hantera externa resultatkrav samtidigt som de ska främja kunskapsutveckling och fostran. Fokus på resultat i skolan har eroderat lärares möjlighet att uppnå agens. Samtidigt har deras forskning visat att lärare ibland övertolkar kraven, då de även när kraven försvagats i styrdokumenten fortsatte att fokusera på resultat. Det kända är ofta mer tryggt än osäkerheten i det okända. Priestley m.fl. beskriver att styrdokumenten ofta uttrycker kravnivåer men ger frihet i valet av metod i hur de uppnås. Det skulle i teorin innebära ett utrymme för läraragens, åtminstone i genomförandet. I praktiken visade det sig dock att lärare ofta anpassade sig efter policy och fokuserade på att samla in underlag för bedömning och klickande i matriser, snarare än att utgå ifrån styrdokumentens syften och mål. Ett svenskt exempel på detta är införandet av betyg i årskurs 6, vilket hade en stor påverkan på lärares bedömningspraktik (Olovsson, 2015). Lärare ökade kontrollen av elevernas prestationer och de hade ett större fokus på prestationer snarare än lärande, då syftet med bedömningarna hade förskjutits till att ge underlag för betygsättning. Lärarnas autonomi i relation till styrdokumenten begränsades och deras bedömningsprocesser blev mer likartade. Även eleverna påverkades på så sätt att de också upplevde en mindre autonomi gentemot styrdokumenten, en ökad disciplin i skolarbetet samtidigt som prestationspressen ökade, vilket många elever uttryckte som negativt.

Lärare försökte hitta en balans mellan det externa ansvarsutkrävandet som uttrycks i läroplanen och en lokal autonomi i klassrumspraktiken (Lidar m.fl., 2020).

Sammanfattningsvis är det flera olika kontextuella villkor som har betydelse för lärares bedömningspraktik. En viktig aspekt är lärares bedömarkompetens. Lärares bakgrund har stor betydelse för lärares bedömningspraktik och deras kompetens. Den bild som framträder i forskningen är att utbildning inte har haft så stor betydelse, vilket ofta kan bero på att kunskaper om bedömning inte har ingått i lärarutbildningar. I stället är det främst lärares egna erfarenheter, preferenser och uppfattningar som har större betydelse, men i början av karriären även kollegor. Vidare har rådande kontextuella villkor, som en performativ diskurs, betydelse för lärares bedömningspraktiker, även om lärare kan övertolka kraven. I avhandlingens diskussion relateras det empiriska materialet till denna forskning.

Bedömningars konsekvenser

Lärares bedömningspraktiker kan även få olika konsekvenser för både elever och samhälle, vilket också är av intresse för denna avhandling. Detta inrymmer även etiska dimensioner. Green m.fl. (2007) beskriver två viktiga etiska principer i relation till bedömningar. Den första är *avoid score pollution*, vilken innebär att lärares bedömningsuppgifter enbart ska bedöma elevers kunskapsutveckling och inte väga in irrelevanta aspekter som ansträngning. Den andra är *do no harm*, vilken innebär att elevens bästa måste vara i centrum. Samtidigt finns det spänningar mellan dessa principer som lärare måste hantera och balansera, där en strikt tillämpning av *no score pollution* kan få negativa konsekvenser för en del elever och tvärtom. Lärare kan också behöva välja mellan olika negativa konsekvenser då en uppgift kan gynna respektive missgynna olika elever på olika sätt. Dessa två principer är intressanta att diskutera i relation till validitet och bedömningars konsekvenser. Om lärare väger in irrelevanta aspekter i bedömningar så påverkas validiteten i både tolkning och användning. Samtidigt behöver lärare beakta negativa konsekvenser av bedömningen för eleverna, vilket också har betydelse för validiteten. I detta avsnitt redovisas forskning om konsekvenser.

En konsekvens är att elever upplever att bedömningsrelaterade frågor ständigt är högst närvarande i skolvardagen (Vogt, 2017). Det handlar såväl om arbete med olika uppgifter som att lärare går igenom och förklarar de komplexa betygskriterierna, vilka eleverna får kämpa för att förstå. Ett stort fokus är på betygsättning och eleverna uttrycker att de får betyg på individuella uppgifter och sällan upplever kreativa lärandesituationer utan att behöva betygssättas. I en intervjustudie skulle elever

uttrycka vad som är en bra lärare och vad som karaktäriserar undervisning som leder till lärande och utveckling. Det visade sig att många av elevernas svar handlade om bedömning och betygsättning som exempel på motsatsen (Hirsh, 2020). Även i denna studie framträdde bilden av bedömning som något konstant närvarande i elevernas skolvardag, vilket ledde till att de kände press och stress för att ständigt behöva prestera. De upplevde att de ständigt bedömdes då lärare ofta uttryckte att allt de gör bedömdes. Detta ledde också till en känsla av att de aldrig får göra fel, vilket ofta hämmade eleverna då de inte vågade fråga eller svara på frågor av risken för att det skulle kunna påverka betygen negativt. Enligt Hirsch tycktes det inte göra någon skillnad om det var summativ eller formativ bedömning, utan det var den ständiga bedömningen som stressade eleverna. Hirsh menar också att den formativa bedömningen tycks mer ha blivit en formell och teknisk process. Detta kan dock problematiseras så tillvida att det skulle kunna vara blandningen av summativ och formativ bedömning som gör att eleverna även känner sig stressade av den formativa bedömningen. Forskning har också visat att när bedömningar används för både betyg och formativa syften så tenderar den formativa effekten av bedömningen att kraftigt minska eller helt utebli (Lauvås m.fl., 2019; Shute, 2008). Således skulle det kunna vara så att det är den performativa diskursen som är grundproblemet och att om eleverna bara bedömdes formativt är det inte säkert att de skulle uppleva samma stress och press.

Vad och hur lärare bedömer påverkar elevernas syn på vad som är viktigt att lära (Brookhart & Durkin, 2003; Löfgren m.fl., 2019; Moss, 2013; Pekrun m.fl., 2023). Bedömningar påverkar elevers motivation men också självförtroende och känslor. Att använda betyg som belöning och straff minskar elevernas motivation att lära sig och skadar själva lärandet. Sättet som lärare presenterar och utformar bedömningar kan påverka elevernas inriktning mot antingen lärandemål eller prestationsmål. Lärare kan stimulera såväl lärandemål som prestationsmål. Till exempel visade en amerikansk fallstudie att tillämpningsuppgifter tenderar att premiera att eleverna blir mer lärande- och målinriktade i stället för prestationsinriktade jämfört med skriftliga prov (Brookhart & Durkin, 2003). Studien visade också att tillämpningsuppgifter bidrar till större självtillit (*self-efficacy*). Andra studier indikerar att bedömning inriktad mot betyg riskerar att leda till att både elever och lärare utvecklar en mer instrumentell syn på skolan (Grettve m.fl., 2014; Löfgren m.fl., 2021; Olovsson, 2015). Det handlar om att bocka av de olika kunskapskraven och det är inte kunskapsinnehållet som fokuseras. Fokus blir mer på betyg än på lärande.

Elevernas uppfattningar om bedömningar kan även påverka deras motivation att lära. Hur bedömningar presenteras och utformas kommu-

nicerar vad som är viktigt och etablerar en kultur i klassrummet, vilket påverkar elevernas motivation och mål (Brookhart & Durkin, 2003). Vidare påverkar bedömningar elevers känslor och vice versa (Löfgren m.fl., 2019; Pekrun m.fl., 2023). Elever som får bra omdömen och höga betyg tenderar att stärka sitt självförtroende och utveckla mer positiva känslor till bedömningar, medan elever som presterar sämre riskerar lägre självförtroende och negativa känslor. Detta kan sedan påverka framtida resultat och förstärka framgång eller motgång. Lärares bedömningspraktiker har således betydelse för både elevernas lärande och välbefinnande.

Rättvis bedömning innebär att alla elever har likvärdiga möjligheter att visa sina kunskaper (McMillan, 2018). Detta kräver transparens om vad som bedöms och hur, men också att det inte förekommer någon bias (Tierney, 2013). En annan aspekt är att eleverna måste ha fått förutsättningar att lära sig det som bedöms. Forskning har visat att lärare ofta använder samma uppgift för alla elever med motiveringen att det är mest rättvist (forskning har också visat att elever ofta resonerar på samma sätt), men detta behöver problematiseras då det i praktiken missgynnar en del elever som har svårt att visa sina kunskaper i vissa redovisningsformer (Tierney, 2013). För en rättvis bedömning behöver lärare ta hänsyn till elevers olika förutsättningar att lära och visa sina kunskaper.

Rättvisan hotas dock om lärare i sina anpassningar av uppgifter även väger in irrelevanta aspekter som kön, etnicitet, socioekonomisk bakgrund eller stereotypiska uppfattningar om beteenden eller förmågor (Tierney, 2013). Forskning har dock visat att sådant förekommer, till exempel vägs ansträngning ofta in i bedömningar (Guskey & Link, 2019; Klapp Ekholm, 2010; Selghed, 2011; Tierney, 2013).

Likvärdig bedömning handlar om att bedömningen ska vara densamma oavsett var i landet den sker, på vilken skola eller vem som bedömer. Likvärdighet är också tydligt uttryckt i styrdokumenten (Skolverket, 2021). Likvärdigheten påverkas negativt av orättvis bedömning men även av andra aspekter. Flera studier pekar på problem med likvärdigheten i bedömningar där olika lärare bedömer samma elevprestation på olika sätt (Black m.fl., 2010; Jølle & Skar, 2020; Skolinspektionen, 2021; Skolverket, 2020; Sundqvist m.fl., 2020; Tengberg m.fl., 2017). En del i detta är det kriterierelaterade betygssystemet, vilket innebär att lärare måste tolka och tillämpa betygskriterierna och det är inte möjligt att alla individer gör samma tolkning och tillämpning (Jönsson, 2023b; Korp, 2011). Ett sätt att öka likvärdigheten är moderande samtal mellan lärare. Forskning visar dock att effekten av sådana samtal inte är entydig (Jönsson & Thornberg, 2014). Dock tycks lärare ofta hamna på samma nivå i bedömningarna men på delvis olika grunder.

Sammanfattningsvis får bedömningar konsekvenser för både elever och samhälle. Elevers motivation, självförtroende och känslor påverkas, vilket i sin tur kan påverka elevernas prestationer i skolan. Vad som bedöms och hur påverkar också elevers lärande. När detta skiljer sig mellan lärare kan det bli problem med rättvisa och likvärdighet, vilket drabbar både elever och samhälle. Det är således av intresse att studera lärares arbete med bedömningar och relatera detta till vilka konsekvenser deras bedömningspraktiker kan ha.

Sammanfattning av tidigare forskning

Sammanfattningsvis finns det inte så mycket forskning om lärares bedömningspraktiker. Den forskning som finns har visat att lärare kan ha olika syften med sina bedömningar, där summativa och formativa syften är centrala. Historiskt har lärare i olika ämnen övervägande använt skriftliga prov med fokus på faktakunskaper. Prov tycks dominera mer i högre åldrar, medan andra former är mer vanliga i lägre åldrar, även om det finns en viss variation och det kan se olika ut i olika länder och mellan olika stadier i skolan. Denna avhandling bidrar med kunskap om hur lärares bedömningspraktiker ser ut i ämnet samhällskunskap i gymnasieskolan. Detta gäller lärares syften med sina uppgifter, vilket innehåll och vilka förmågor som testas samt med vilka metoder.

Forskning har pekat på olika faktorer som har betydelse för lärares bedömningspraktik. Flera studier har visat att lärare ofta saknar utbildning om bedömning, medan tidigare erfarenheter tycks ha större betydelse. Även andra aspekter som lärares preferenser och känslor kan påverka. Den performativa diskursen har också visat sig ha betydelse för lärares agens, samtidigt som lärare kan övertolka kraven. Ämnets karaktär tycks ha betydelse för bedömningar i olika ämnen, åtminstone vid betygsättning. Hur karaktären påverkar enskilda uppgifter är mer outforskat. Denna avhandling bidrar med kunskap om faktorer som påverkar lärares beslut om bedömningar, liksom om ämneskaraktärens betydelse.

Lärares bedömningspraktiker får också konsekvenser, där forskning har pekat på att elevers lärande, motivation, känslor och självtillit påverkas. Vidare får praktiken konsekvenser för likvärdighet rättvisa och validitet och forskningen pekar på en hel del problem gällande detta och avhandlingens resultat relateras till denna forskning om konsekvenser.

5 Metod

Avhandlingen handlar om att studera lärares bedömningspraktik gällande syften med bedömningarna, vilket innehåll och vilka förmågor som bedöms samt med vilka metoder. Likaså studeras faktorer som kan påverka dessa praktiker. För att besvara detta valde jag semistrukturerade intervjuer. I delstudie I och III kombinerades dessa med insamlade artefakter i form av lärares bedömningsuppgifter: prov i delstudie I och alla bedömningsuppgifter (vilka fanns i skriftlig form) i delstudie III. Uppgifterna analyserades utifrån vilket innehåll och vilka förmågor som testades samt vilka metoder som användes för att testa detta. Genom att analysera lärarnas bedömningsuppgifter synliggörs deras faktiska praktik i form av vilka uppgifter de använder, liksom vilka kunskaper och förmågor dessa mäter. Samtidigt ger dessa inte hela bilden och därför kompletterades de med intervjuer, vilka också behövdes för att kunna analysera olika faktorer som kan påverka lärares bedömningspraktiker.

Datainsamling

Två olika datainsamlingar gjordes, en för den första delstudien (licentiatuppsatsen) och en för de andra två (artiklarna). I båda datainsamlingarna genomfördes semistrukturerade intervjuer men med olika inriktning och frågor. Urvalet gjordes också på lite olika sätt.

Den första datainsamlingen

I den första datainsamlingen genomfördes först en förstudie genom att 135 lärare, vilka utgjorde samtliga lärare i samhällskunskap i ett medelstort län i Sverige, ombads att skicka två prov (ett om hur Sverige styrs och ett om samhällsekonomi). Av dessa skickade 21 lärare in prov. Proven analyserades med hjälp av Blooms reviderade taxonomi. Utifrån detta delades lärarna in i tre kategorier beroende på hur hög andel minnesfrågor det var i proven (<50%, 50–80% samt >80%). På detta sätt skapades två ”extremer” samt en mellangrupp. Två lärare från varje kategori valdes ut, sammanlagt sex, och kontaktades för en uppföljande intervju. En av dessa avböjde, varpå en annan lärare i samma kategori kontaktades och denne tackade ja. Det var två kvinnor och fyra män i olika åldrar. De arbetade på olika skolor i både större och mindre kommuner och både friskolor och kommunala skolor. De undervisade också på både yrkesförberedande och studieförberedande program (vilket idag heter yrkesprogram respektive högskoleförberedande program). På detta sätt gjordes ett strategiskt urval med så stor spridning som möjligt (Trost &

Hultåker, 2016) för att möjliggöra att täcka en stor del av den variation som kan finnas i lärares bedömningspraktiker och deras tankar och uppfattningar om sina praktiker.

Intervjuerna genomfördes hösten 2009 på lärarnas arbetsplatser. Tanken var att det skulle vara en plats som var bekant för lärarna där de kunde känna sig bekväma och avslappnade (Kvale & Brinkmann, 2014). Intervjuerna varade i genomsnitt 70 minuter (mellan 52 och 79 minuter). De semistrukturerade intervjuerna skedde utifrån en intervjuguide (se bilaga 2 i delstudie I) och vid intervjuerna fanns även lärarnas inskickade prov med. Frågorna handlade om hur lärarna tolkade betygskriterierna, prov generellt (syftet med dem, vad de kan användas till), specifika frågor om de inskickade proven (vad de avsåg mäta och kontexten kring dem) samt om olika faktorer som påverkar lärarnas arbete med prov. Intervjuerna spelades in med diktafon och transkriberades i stort sett ordagrant men anpassat efter skriftspråkets förutsättningar utan att förlora innehållets mening (Kvale & Brinkmann, 2014).

Den andra datainsamlingen

I den andra datainsamlingen, för delstudierna II och III, skickades en förfrågan till 108 samhällskunskapslärare i 17 olika kommuner av varierande storlek i tre olika län. De ombads skicka in en komplett uppsättning uppgifter som de använt för den första kursen i samhällskunskap (1b/1a1) på gymnasiet. Det var 16 lärare som skickade in material, vilka tillsammans hade en stor spridning i de uppgifter de använde. Detta bedömde jag som tillräckligt och intressant, då ett material med stor spridning kan ge bra underlag för en tematisk analys (Braun & Clarke, 2022). Därför kontaktade jag dessa 16 lärare för en uppföljande intervju, varav 13 tackade ja. Det var 7 män och 6 kvinnor, varav alla utom en hade svensk bakgrund. Tio arbetade i kommunala skolor och tre på friskolor. Två av lärarna arbetade på samma skola. Skolorna varierade i storlek och låg på både större och mindre orter i olika regioner. Alla var erfarna lärare med lärarexamen från mellan 1988 och 2014. Dock var en inte behörig i samhällskunskap utan i andra ämnen. Det innebar att ingen av dem hade läst en kurs i bedömning under sin lärarutbildning, då en sådan kurs blev obligatorisk först för studenter som började sin utbildning höstterminen 2011. Därigenom ger detta ett strategiskt urval med en stor spridning (Trost & Hultåker, 2016), vilket gör det möjligt att finna en stor del av variationen i lärares bedömningspraktiker, liksom deras resonemang om detta.

Intervjuerna genomfördes våren 2019 på lärarnas arbetsplatser med undantag av en intervju som läraren valde att förlägga till ett café. De varade i genomsnitt 80 minuter (mellan 46 och 106 minuter). En intervjuguide användes (se bilaga i delstudie III) med breda och öppna frågor

om lärarnas bedömningspraktiker (till exempel beskriva sin bedömningspraktik samt hur de resonerar om val av olika bedömningsformer). Breda och öppna frågor valdes för att möjliggöra för lärarna att kunna tala relativt fritt om sin bedömningspraktik. Uppföljningsfrågor kunde handla om faktorer som påverkade deras metodval, relationer mellan metoder och innehåll, liksom mellan olika kunskapsformer och metoder. Specifika frågor om deras inskickade uppgifter ställdes också. Även dessa intervjuer spelades in med diktafon och transkriberades så ordagrant som möjligt.

Analyser av data

I delstudie I används en form av kvalitativ textanalys (Boréus & Bergström, 2018) där de inskickade proven tillsammans med lärarnas utsagor om de enskilda frågorna analyserades (flera gånger) med hjälp av Blooms reviderade taxonomi (Anderson m.fl., 2001) för att identifiera vilka kunskapsformer som de olika provfrågorna kan testa. Taxonomin, och analysprocessen förklaras mer utförligt i den teoretiska bakgrunden i delstudie I. Lärarnas utsagor om betygen analyserades också med hjälp av taxonomin, men också utifrån att hitta uttryck för ramfaktorer som kunde påverka lärarnas arbete med bedömningar, liksom uttryck för ett frirum i relation till dessa ramar.

Tematisk analys valdes för att identifiera olika syften med lärares bedömningar i intervjuerna i delstudie II (Braun & Clarke, 2022). För att finna teman om vilket innehåll och förmågor som testas, samt med vilka metoder använde jag även i delstudie III tematisk analys av både inskickade uppgifter och intervjuerna. De tematiska analyserna genomfördes med stöd av datorprogrammet Nvivo. Intervjuerna analyserades också utifrån hur lärarna resonerade om sina beslut om bedömningar och om faktorer som kunde påverka deras bedömningspraktiker.

Den tematiska analysprocessen (Braun & Clarke, 2022) innebar att jag flera gånger både lyssnade och läste igenom intervjuerna och lärarnas uppgifter för att bli bekant med materialet. Sedan kodades materialet induktivt, nära empiriska data. I kodningen användes både semantiska (utifrån de explicita och konkreta uttrycken i utsagorna) och latenta koder (underliggande innebörder i utsagorna). I delstudie II sökte jag efter uttryck som relaterade till lärares syften med sina bedömningsuppgifter. I delstudie III letade jag efter utsagor som handlade om vilka bedömningsmetoder lärarna använder, vilket kunskapsinnehåll och vilka förmågor som de testar med sina uppgifter, liksom uttryck om relationer mellan dessa. Vidare sökte jag efter resonemang om faktorer som kan påverka lärarnas beslut om val av innehåll, förmågor och metoder. Även

lärarnas inskickade uppgifter kodades utifrån vilket innehåll och vilka förmågor som testades samt vilka metoder som användes.

Utifrån detta genererades initiala teman och underteman och koderna sorterades i dessa teman. Alla initiala teman och underteman inkluderades oberoende av hur frekvent de förekom. Teman och underteman granskades och reviderades för att få fram klara definitioner och benämningar, vilket till sist genererade olika huvudteman med flera underteman i respektive delstudie.

I delstudie III sorterades de induktiva temana i den tematiska analysen av vilket innehåll som testades i relation till det centrala innehållet i kursplanen för Samhällskunskap 1b (Skolverket, 2022c). I studien användes också Blooms reviderade taxonomi (Anderson m.fl., 2001) för att kategorisera teman om vilka förmågor som lärarnas uppgifter testar. En förenkling gjordes såtillvida att kategorierna faktakunskap och konceptuell kunskap i kunskapsdimensionen slogs ihop. Båda dessa är en form av faktakunskap, med skillnaden att faktakunskap är mer konkreta fakta och konceptuell kunskap mer abstrakt fakta. Bedömningen var dock att det inte var meningsfullt att skilja dessa åt i analysen, då båda kategorierna är en form av faktakunskap, men med olika abstraktionsnivå. När det i sammanfattningen av studien refereras till faktakunskaper avser det således både fakta- och konceptuell kunskap

Forskningsetiska överväganden

Lärarna i båda studierna fick själva välja om de ville skicka in material och ställa upp för intervjuer (missivbrev finns i bilagorna 3 och 4). De informerades både i förväg och vid intervjuerna om studiernas syften, om frivilligheten i att delta och om att de när som helst kunde avstå från att delta. De informerades även om att de skulle anonymiseras i studierna och därför är de namn som används i avhandlingen fiktiva. De gav sitt samtycke till att delta, dels genom att tacka ja via email, dels muntligt i samband med intervjuerna (vilket spelades in). Därmed kan de forskningsetiska principerna anses vara uppfyllda (Vetenskapsrådet, 2017). Insamlade data, både intervjuer och lärarnas uppgifter, förvaras digitalt på Linköpings universitets Onedrive. Eftersom materialet inte innehåller några integritetskänsliga data bedöms detta vara tillräckligt.

Metoddiskussion

Jag valde semistrukturerade intervjuer som metod för att undersöka faktorer som påverkar lärarnas beslut om bedömningar. De ger också underlag för lärarnas resonemang om sin bedömningspraktik. Semistrukturerade intervjuer är en flexibel metod som möjliggör för intervjuaren

att anpassa intervjun efter informanternas svar, vilket kan leda till mer informativa svar (Bryman, 2016; Kvale & Brinkmann, 2014). Det möjliggör därmed en djupare förståelse av personers tankar och uppfattningar. I semistrukturerade intervjuer används intervjuguider, vilka utformas utifrån olika teman, med möjliga följdfrågor. Detta ger en flexibilitet som kan möjliggöra att informanterna kan svara relativt fritt och inte styrs av intervjuaren. Det ger också en mer samtalsliknande och förtroendeskapande situation som kan öka förutsättningen för att informanterna känner sig avslappnade och ger ärliga svar. Samtidigt ger intervjuguiderna en struktur som gör att det går att jämföra informanternas olika svar.

En epistemologisk utgångspunkt för intervjuer är att det lärare säger är ett slags verklighet som forskare kan få kunskap om genom intervjuer (Kvale & Brinkmann, 2014). I intervjuer konstrueras kunskap socialt i interaktionen som sker mellan informant och intervjuare. Denna konstruktion påverkas av intervjuarens perspektiv och värderingar och intervjuer innebär också en tolkning av meningen i det som sägs. Därför blir det också viktigt med kontexten och uttolkarens förkunskaper. Kontexten är interpersonell, där det finns ett samband mellan kontexten och meningen i informantens utsagor. I både konstruktionen av kunskap och tolkandet är språket viktigt. Det är inte så att språket och kunskapen kopierar verkligheten, utan de är redskap för att hantera denna.

Båda urvalen bygger på att deltagarna själva valt att delta genom att skicka in material och att ställa upp på intervjuer. Det innebär en risk för skevhet i urvalet (Teorell & Svensson, 2007). Till exempel finns en risk att de som känt att deras bedömningsuppgifter inte är tillräckligt bra låtit bli att delta, eller att lärarna skickat sina ”bästa” uppgifter och dessa inte är representativa för deras bedömningspraktik. Detta ser jag dock inte som ett problem då syftet inte är att kvantifiera resultaten, utan att finna kvalitativa teman. Eftersom det i båda datainsamlingarna är en relativt stor spridning mellan lärarnas bakgrunder och bedömningspraktiker, är min bedömning att det täcker in en tillräcklig bredd av lärares bedömningspraktiker.

I en kvalitativ studie är det viktigt med en perspektivmedvetenhet och att redovisa sin förförståelse (Larsson, 2005), vilket även betonas i tematisk analys (Braun & Clarke, 2022). Det är också av betydelse utifrån perspektivet att kunskapen konstrueras i interaktionen mellan informant och intervjuare, där intervjuarens perspektiv och värderingar påverkar (Kvale & Brinkmann, 2014). Forskarens perspektiv och värderingar kan även påverka tolkningen av lärarnas uppgifter. Jag är utbildad och tidigare verksam gymnasielärare i samhällskunskap, vilket kan vara både en fördel och en nackdel. En fördel är att egna kunskaper och erfarenheter av ämnet, liksom av arbetet med bedömningar, ger en förståelse för det som informanterna säger och deras inlämnade uppgifter. En

kunnig intervjuare lyfts fram som positivt (Kvale & Brinkmann, 2014). En intervjuare som inte är expert skulle dock kunna inta en roll som intresserad lärande där i stället informanten ges rollen som expert (Braun & Clarke, 2013). Samtidigt ger goda kunskaper inom området större möjligheter till att ställa relevanta följdfrågor under intervjun. En nackdel är att de egna erfarenheterna skulle kunna innebära en ”hemmablindhet” och att författarens perspektiv påverkat tolkningen av analyserna i en viss riktning eller att sådant som tas förgivet kan förbli osynligt. Ytterligare en fördel med egen erfarenhet har varit intervjusituationerna, där informanterna har kunnat känna sig trygga med en ämneskollega (Kvale & Brinkmann, 2014). En begränsning med intervjuer generellt är att det finns en risk för sociala önskvärdhetseffekter, det vill säga att informanterna svarar som de tror att intervjuaren önskar (David & Sutton, 2016). Andra felkällor kan vara bristande minne och perception. Det finns således en risk att det informanterna säger inte behöver stämma med vad de faktiskt gör eller tänker. Avhandlingens olika delar har behandlats i seminarier, licentiatuppsatsen har granskats vid en framläggning och artiklarna har granskats genom tidskrifternas peer-review-förfarande. Därigenom har resultaten prövats intersubjektivt, vilket stärker delstudiernas och avhandlingens trovärdighet (Boréus & Bergström, 2018).

En annan kvalitet i kvalitativa studier är att de möjliggör innebördsrikedom (Braun & Clarke, 2022; Larsson, 2005). I avhandlingens delstudier har det eftersträvats att ge en så fyllig beskrivning av empirin som möjligt. Samtidigt behövs en god struktur med en överskådlighet och en reduktion av komplexitet (Larsson, 2005). Det finns en spänning mellan dessa krav, vilka behöver balanseras. I delstudie I är därför redovisningen av empirin relativt omfattande, både gällande de prov som analyseras och intervjuerna med lärarna. I delstudie II och III har också ambitionen varit att vara generös med citat från informanterna. Det begränsade artikelformatet innebär dock att det inte är lika många citat som i licentiatuppsatsen, vars format är mer omfattande. Att redovisa så mycket empiri som möjligt bidrar till en större transparens och ökar trovärdigheten i att studiens resultat har en empirisk förankring (Larsson, 2005).

I avhandlingen används ibland uttryck som ”alla lärare”, ”några lärare”, ”påverkar ofta” eller har ”liten betydelse”. Uttrycken grundas på hur många av delstudiernas lärare som uttrycker något, men också om det är något som de ofta återkommer till, eller betonar i sina utsagor. Detta är dock en kvalitativ studie vilket innebär att de inte ska tolkas som kvantitativa uttryck med generaliserbara anspråk, utan de är endast indikationer och ytterligare forskning behövs för mer generaliserbara resultat. I sammanfattningarna och i diskussionen skrivs lärarna i bestämd form när det avser lärarna i studierna och lärare i obestämd form när det syftar på lärare i allmänhet.

6 Sammanfattning av delstudierna

Avhandlingen består av tre delstudier, en licentiatuppsats och två artiklar, vilka sammanfattas nedan.

Delstudie I (licentiatuppsats) – Vad kommer på provet?

I denna studie undersöktes provpraktik hos gymnasielärare i samhällskunskap, med särskilt fokus på vilka kunskapsformer som lärarna testar i sina prov. Två prov från vardera av de sex lärarna i studien analyserades med hjälp av Blooms reviderade taxonomi. Dessa analyser kompletterades med semistrukturerade intervjuer (Kvale & Brinkmann, 2014) om proven, men också om lärarnas provpraktik i allmänhet, då studien också berörde lärarnas syften med sina prov, vilka möjligheter och begränsningar de ser med prov och vad som kan påverka utformandet av prov. Mot bakgrund av reformerna 1994, med bland annat införandet av kriterierelaterade betyg, tog studien också upp hur lärarna uttrycker skillnader vad gäller kunskapsformer mellan olika betygsnivåer och analyserade hur deras provpraktik förhåller sig i relation till detta.

Lärarna i studien uttryckte att de använder prov som underlag för betyg, det vill säga i summativt syfte. De menade att prov kan användas för att testa de flesta kunskapsformer och därför ansåg de att prov var en bra bedömningsform. Ett undantag var *procedurkunskaper*, vilka främst handlade om den samhällsvetenskapliga metoden som de ansåg var bättre att testa med andra uppgifter som uppsatser.

Tidigare studier (Mattson, 1989; Svingby, 1998) har visat att lärares prov främst är faktaorienterade, det vill säga att de testar att *minnas*. Lärarnas prov i denna studie visade sig i analyserna ha en stor variation i vilka kunskapsformer som testas. Även om alla sex lärarna hade en relativt hög andel minnesfrågor, mellan 50 och 90 procent, så testades också andra kunskapsformer i deras prov, främst att *förstå*, men även att *analysera* och *värdera*. Sammantaget fanns frågor som testade alla kognitiva förmågor i taxonomin, även om det var mindre vanligt med frågor som testade att *tillämpa* eller att *skapa*. Samtidigt var det en variation i testade kunskapsformer mellan lärarnas prov, där en del av lärarna hade en större variation än andra. Det innebar att eleverna utifrån proven testades på olika former av kunskap. Eftersom samtliga lärare också använde proven som underlag för betyg och gav betygsomdömen på alla betygsnivåer, fanns det också en risk att eleverna fick betyg på olika grunder. Till exempel gick det i några prov att få alla betygsnivåer genom att bara använda minneskunskaper. Tidigare studier (Black m.fl., 2010;

Skolinspektionen, 2010; Skolverket, 2009) har visat på problem med likvärdigheten i lärares bedömningar, vilket också denna studie indikerade.

Däremot uttryckte sig lärarna relativt lika när de resonerade om skillnaderna mellan de olika betygen och vad betygskriterierna innebär. De tycktes således ha en samsyn om tolkningen av betygskriterierna men deras provpraktik skilde sig åt. Lärarna var överens om att betygskriterierna för Godkänt (G) främst handlade om minneskunskaper, medan de högre betygen, Väl godkänt (VG) och Mycket väl godkänt (MVG) mer handlade om att *analysera* och *värdera*. En liten skillnad, utifrån studiens analyser, var att lärarna också relaterade att förstå till de högre betygen, medan betygskriterierna snarare relaterade förståelse till betyget Godkänt. Lärarnas prov stämde således inte alltid helt överens med de betygskriterier som de var tänkta att testa. Det innebär att det i likhet med andra studier fanns en diskrepans mellan lärarnas övertygelser och deras provpraktik (Black m.fl., 2010). Detta gällde både om proven jämfördes med själva betygskriterierna och om de jämfördes med lärarnas egna beskrivningar av betygen. Inom gruppen fanns det dock en variation och diskrepansen mellan betygssyn och provpraktik var större för en del lärare än andra. Detta innebar att det förutom problem med likvärdigheten också fanns problem med validiteten i lärarnas prov.

Olika ramfaktorer (Imsen, 1999; Lundgren, 1989) har betydelse för lärares utformning av prov och dessa ramar kan bidra till att förklara denna diskrepans. Lärarna lyfte speciellt fram några ramar i intervjuerna. Styrdokumenten, vilka är pedagogiska ramar, har enligt lärarna stor betydelse. Lärarna uttryckte dock ett stort frirum i relation till styrdokumenten. Till exempel tycktes lärarna inte förhålla sig till kursmålen i någon högre grad, utan det var främst betygskriterierna som påverkade. I tolkningen och tillämpningen av betygskriterierna kompromissade också lärarna i relation till andra ramar som arbetstid och lektionstid, vilka är administrativa ramar, samt elevernas önskemål, vilka är organisationsrelaterade ramar.

Samtidigt tycktes lärarna uppfatta frirummet i relation till de administrativa ramarna olika. Det var främst lärarnas arbetstid och schemat som begränsade lärarnas arbete med prov. Lärarna behövde tid att förbereda och tid att bedöma elevsvaren, samtidigt som eleverna behövde tid för att skriva proven. Om lärarna ska kunna ha prov med en stor andel frågor i de mer komplexa kognitiva kunskapsformerna underlättas det således av att lärarna har tid för detta liksom att schemat kan påverkas. Några av lärarna i studien fann ett frirum genom att till exempel kunna byta lektioner med en kollega för att skapa längre skrivtid för eleverna. Skolornas sätt att hantera lärares arbetstid och schemaläggning har således betydelse för lärarnas möjligheter att göra valida prov i relation till betygskriterierna.

Elevernas önskemål, vilka är organisationsrelaterade ramar, har också betydelse för lärarnas provpraktik. Lärarna uttryckte att eleverna ofta föredrar skriftliga prov framför andra examinationsformer och de ville också helst ha prov med minnesfrågor, där det finns tydliga rätt och fel. Det tycks finnas en kultur bland eleverna att betona minneskunskaper. Detta tog lärarna viss hänsyn till, både gällande antalet prov och vad de testade, även om de inte anpassade sig helt. Resursrelaterade och elevrelaterade ramar tycktes däremot inte ha någon större betydelse för lärarnas provpraktik.

I motsats till samhällskunskapsdidaktisk litteratur (Arevik & Hartzell, 2015; Lindqvist & Hyltegren, 1995; Svingby, 1989) menade lärarna i studien att ämnet samhällskunskap inte har någon särskild karaktär som påverkar utformandet av prov. De trodde inte att prov i samhällskunskap skiljer sig mot prov i andra ämnen, dock med undantag av matematik som några lärare pekar ut som annorlunda. Lärarna ansåg också att kunskapsformerna i betygskriterierna inte skiljer sig mellan ämnena, utan att det handlar om universella förmågor.

Lärarna hade under intervjuerna svårt att tala om olika kunskapsformer, vilket tyder på en brist på ett gemensamt ämnesdidaktiskt språk (Schüllerqvist & Osbeck, 2009). Samtidigt behöver det, enligt Uljens (1997), finnas en gemensam förståelse för de begrepp som används. Detta behövs för en yrkesprofessionell kommunikation och möjliggör lärares egen reflektion över sin yrkespraktik. Lärarna använde dock ibland samma kognitiva begrepp, till exempel *analysera*, men de tycks inte alltid mena samma sak. Blooms reviderade taxonomi visade sig användbar i studien för att resonera om olika kunskapsformer och skulle kunna vara ett verktyg för att utveckla ett ämnesdidaktiskt språk om kunskapsformer. En viktig aspekt i taxonomin är att för att en uppgift ska testa andra kognitiva förmågor än att *minnas* så behöver den innehålla något kvalitativt nytt för eleverna, så att det inte räcker med att komma ihåg något från läromedel eller undervisning. I studien visade det sig att alla lärare inte är medvetna om detta, utan de uttryckte ibland att frågor var tänkta att testa en högre kognitiv förmåga men konstaterade samtidigt att det räckte med att eleverna kom ihåg till exempel en diskussion från lektionerna. Lärare behöver således problematisera mer om detta.

Sammanfattningsvis är studien i linje med tidigare forskning och rapporter som pekar på problem med likvärdighet och validitet i lärares bedömningar (Black m.fl., 2010; Skolinspektionen, 2010; Skolverket, 2009). Även om lärarna i studien tycktes ha samsyn i tolkningen av kriterierna, skilde de sig åt i tillämpningen av dessa, åtminstone i sina prov. Många lärare skulle dock kunna göra mer valida prov, men förutsättningen är att de får tid till detta, både att förbereda och att bedöma, samt att de har möjlighet att ha tillräckligt långa lektioner så att eleverna har

tid att svara på frågorna. Samtidigt behöver lärare problematisera till exempel skillnaden mellan att *minnas* och högre kognitiva processer. Lärarutbildningen skulle kunna förbereda blivande lärare för en sådan reflektion.

Delstudie I visade en del intressanta resultat. Samtidigt studerades bara lärares prov och det vore också intressant att studera alla uppgifter som lärare använder för att få en bredare bild av samhällskunskapslärares bedömningspraktik. Detta ledde till de andra två delstudierna.

Delstudie II (artikel) – Balancing diverse purposes in Civics teachers' assessment decision making

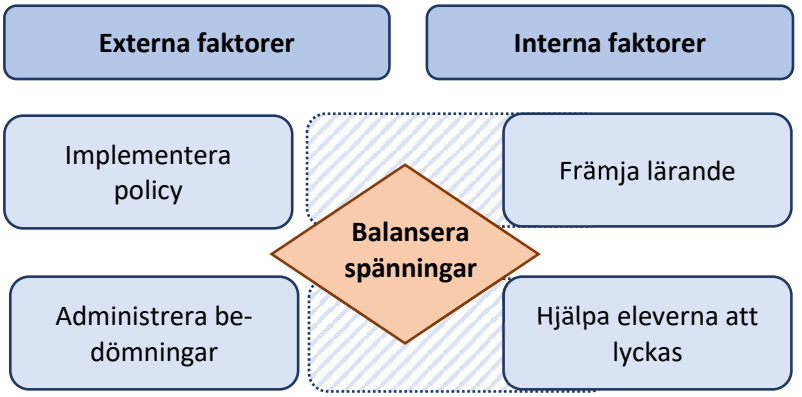
Syftet med denna studie var att beskriva de olika syften med bedömningar som svenska samhällskunskapslärare på gymnasiet uttrycker. Studien bidrar till en förståelse av hur olika interna och externa faktorer påverkar lärares beslut om bedömningar men också till den pågående debatten om validitet i lärares bedömningar. Studien bygger på semistrukturerade intervjuer (Bryman, 2016) med tretton lärare med olika bakgrund och i olika kontexter. Intervjuerna fokuserade på lärares bedömningspraktik gällande deras val av vad som bedöms och hur, samt grunderna för dessa val. Tematisk analys (Braun & Clarke, 2006, 2013) användes för att hitta kvalitativa kategorier av lärarnas syften.

I studien användes en modell om beslutsfattande för bedömningar om olika interna och externa faktorer som påverkar lärares beslutsfattande och därmed deras bedömningspraktik (McMillan, 2003). Modellen visar också att det finns spänningar mellan olika faktorer.

I en bedömningsprocess fattar lärare mängder av beslut: de väljer ut vilket innehåll och vilka förmågor som ska bedömas, med vilka metoder samt hur resultaten ska bedömas och användas (McMillan, 2018). Lärarnas syften med bedömningarna är viktiga för deras val av innehåll och vilka metoder som är mest relevanta (Newton, 2017). Resultaten visade att lärarna hade många och olika syften med sina bedömningar, vilket är i linje med tidigare forskning (Harris & Brown, 2009; McMillan, 2001). Dessa olika syften kunde delas in i fyra olika huvudteman, vilka relaterar till interna och externa faktorer (figur 6).

Det första temat, *implementera policy*, relaterar till externa faktorer. Lärarna uttryckte syften att uppfylla krav i styrdokumenten, till viss del i läroplanen, men främst i ämnesplanen och särskilt då kunskapskraven. Att ha underlag för betygsättning var ett starkt syfte för lärarna. De ville också att bedömningarna ska vara transparenta och därför är det viktigt att kommunicera och tydliggöra mål och kunskapskrav så att eleverna känner till grunderna för bedömningen. Lärarna talade också om att bedömningar ska vara rättvisa och ville därför testa varje kunskapskrav

Figur 6. *Balansera spänningar mellan olika syften med bedömningar (Jansson & Löfgren 2022).*



minst två gånger. Detta var dock ibland svårt; lärarna uttryckte en frustration över vissa kunskapskrav som handlade om ett specifikt ämnesinnehåll, vilka lärarna tyckte var svåra att hinna testa mer än en gång.

Det andra temat är syften som handlar om att *administrera bedömningar*, vilka också är externa faktorer. Lärarna uttryckte syften om att bedömningar måste rymmas inom arbetstiden, inom kurstiden och i schemat. Andra administrativa syften handlade om att de ville återanvända uppgifter, dels för att de fungerar och för att de kände sig trygga med dem, dels för att spara tid. Samtidigt uttryckte en del att de förnyar sina uppgifter för att utmana sig själva. Lärarna hade också som syfte att förebygga fusk.

Syften att *främja lärande* är det tredje temat, vilket främst relaterar till interna faktorer. Även om det finns formuleringar i styrdokumenten om lärande, uttryckte lärarna det främst som en egen inre vilja. Att främja lärande handlar mycket om formativ bedömning, där lärarna främst uttryckte syften om att ge återkoppling, men i viss grad också om kamratbedömning och självbedömning. Lärarnas syfte var också att främja djupinlärning och lärarna ville att eleverna ska förberedas för livet, för kommande gymnasiekurser och för högre studier. De hade också ett ämnesspecifikt syfte om att eleverna ska lära sig om dagsaktuella händelser, vilket de såg som centralt i samhällskunskap.

Det sista temat är syften att *hjälpa eleverna att lyckas*. Även detta syfte relaterar främst till interna faktorer, även om det finns liknande mål i läroplanen, då det främst är en inre drivkraft hos lärarna. Lärarna uttryckte syften om att ge eleverna inflytande över bedömningar för att öka chansen att lyckas, genom att välja tidpunkt och format för bedömningarna. Dessa val kunde göras i helklass, grupp eller individuellt. Lär-

arna gjorde också anpassningar utifrån elevernas program, klassens förutsättningar och individuella behov med syftet att hjälpa eleverna att lyckas. De ville också ge eleverna flera tillfällen att visa sina kunskaper. Ett annat syfte var att utforma bedömningar så att de motiverar eleverna, vilket ökar chansen att lyckas. Till sist främjas elevernas framgång genom att lärarna ville minska elevernas stress och arbetsbörda.

Lärare behöver balansera dessa olika syften och de spänningar som finns mellan dem, vilket också tidigare studier visat (Harris & Brown, 2009; McMillan, 2001). Liksom tidigare studier (Brookhart, 2009; Gioka, 2009), visade studien på spänningar mellan formativ och summativ bedömning. Dessa spänningar förstärks av de svenska styrdokumentens betoning på betygsättning, samtidigt som det finns en diskurs som förespråkar formativ bedömning (Wahlström & Sundberg, 2015). Lärarna i studien gav ofta betygsomdömen eller använde kunskapskravens formuleringar i sin återkoppling, vilket riskerar att eleverna inte tar till sig formativ återkoppling (Shute, 2008). En annan spänning är mellan lärarnas syften att främja lärande liksom att hjälpa eleverna att lyckas och syften att administrera bedömningar. Det kan vara tidskrävande att skriva formativ återkoppling liksom att göra anpassningar, vilket kan vara svårt att hantera inom arbetstiden (Lumadi, 2013; Samuelsson m.fl., 2018). Syftet att ge eleverna fler tillfällen kan också vara svårt att rymma inom arbetstiden och kurstiden.

Ämnens karaktär påverkar lärares beslut om bedömningar (Prøitz, 2013). Lärarna uttryckte syften att uppfylla kraven i ämnesplanen. Framför allt är det kunskapskraven som styr men dessa kan vara problematiska då några av dem har ett specifikt faktainnehåll. Det uppstår spänningar med syftet om rättvisa. Kunskapskraven uttrycker också att eleverna måste kunna visa sina kunskaper både muntligt och skriftligt men en del elever har svårt i den ena eller den andra formen. Detta leder till spänningar mellan att *implementera policy* och att *hjälpa eleverna att lyckas*. Aktualisering är en del av ämnets karaktär, vilket kräver att uppgifter uppdateras. Det kan också ta en del tid och skapar spänningar med lärares arbetstid. Ämnet samhällskunskap brukar beskrivas som öppet, flexibelt och med svag inramning (Bernstein, 2003; McMillan, 2001). I Sverige kan denna bild nyanseras, då ämnesplanen har en del ganska snäva formuleringar i form av de specifika kunskapskraven. Tidigare studier har beskrivit att ämnesplanen snarare fokuserar på generella förmågor än på ämnesspecifika (Wahlström m.fl., 2020). Studien visade dock att lärarna uppfattade dessa förmågor i relation till ämnet då de till exempel såg analysera som annorlunda i samhällskunskap än i andra ämnen.

Validitet i lärares bedömningar handlar om att lärares slutsatser och användning av bedömningar behöver vara relevanta i relation till be-

dömningens syfte (Kane, 2016; Messick, 1989). Ett problem är dock att teorierna om validitet utvecklats i andra sammanhang än skolan och kraven på validitet är inte alltid tillämpliga i skolmiljön (Bonner, 2013; Cizek, 2009b; Kane & Wools, 2020). Det kompliceras också av att lärare har många och olika syften (Newton, 2017). En bedömning kan inte vara helt valid för alla olika syften och därför behöver lärare kompromissa mellan sina olika syften. För att validitet ska vara meningsfullt för lärare i skolan behövs en förenklad valideringsprocess som är hanterbar för lärare (Bonner, 2013). En modifierad och förenklad variant av Kanes (1992, 2016) argumentsbaserade tillvägagångssätt för validering skulle kunna användas. Lärare bör identifiera sina syften och prioritera mellan dessa och valideringen görs med fokus på bedömningens huvudsyften. Samtidigt bör läraren hålla begränsningarna i validitet för andra syften i minnet (Kane, 2016; Newton, 2017). Olika syften ställer olika krav på validitet. Till exempel är kraven högre för betygsättande syften än för formativa syften, vilket också behöver beaktas. För hög validitet bör olika förmågor testas vid flera tillfällen, vilket dock inte alltid är möjligt, till exempel för de innehållsspecifika kunskapskraven i samhällskunskap. I praktiken behöver lärarna göra en mer ytlig bedömning av validiteten i sina bedömningar för att det ska vara praktiskt hanterbart.

Sammanfattningsvis har svenska samhällskunskapslärare på gymnasiet många och olika syften med sina bedömningar. I sitt dagliga bedömningsarbete behöver de balansera dessa syften och spänningarna emellan dem i sina beslut om olika bedömningsuppgifter, vilket formar deras bedömningspraktiker. De behöver validera sina bedömningar med en förenklad valideringsprocess som fokuserar på huvudsyftena med respektive bedömningsuppgift. Med en större förståelse för komplexiteten i lärares bedömningsprocess, med många olika syften, är lärare förhoppningsvis bättre rustade att fatta ändamålsenliga beslut om sina bedömningar.

Delstudie III (artikel) – Civics teachers' assessment practices in Swedish upper secondary schools. A qualitative study.

Syftet med denna delstudie var att undersöka samhällskunskapslärares beslut om vilket innehåll och vilka förmågor som ska bedömas och med vilka metoder, liksom vad som kan påverka dessa beslut. Tretton lärares inskickade bedömningsuppgifter analyserades tematiskt, kompletterat med semistrukturerade intervjuer om deras bedömningspraktik och om de inskickade uppgifterna. Teorin om läraragens användes för att diskutera olika påverkande faktorer.

Samhällskunskapslärarna uttryckte att de inte hade ambitionen att bedöma alla delar av det centrala innehållet, även om allt togs upp i undervisningen. Tidigare studier har visat att även om ämnet anses öppet, flexibelt och med svag inramning så tycks det finnas en ämnestradition med en ganska stark kanon (Odenstad, 2011; Olsson, 2016; Öberg & Bäckström, 2021). Detta bekräftades i studien av både inskickade uppgifter och lärarnas utsagor, vilka visade att det var mer vanligt att statsvetenskap, särskilt om hur Sverige styrs, och samhällsekonomi bedömdes, medan andra områden var mindre vanliga. Några områden som inte tidigare studier har lyft fram, var att lärarna i studien också ofta bedömde mänskliga rättigheter, privatekonomi och källkritik. De två förstnämnda förklarades av att det fanns särskilda kunskapskrav för detta innehåll och lärarna kände att de måste bedöma detta. Lärarna uttryckte att de ser källkritik, liksom aktualisering, som en del av ämnets karaktär. Lärarna relaterade därför ofta uppgifternas innehåll till aktuella händelser.

Tidigare studier har visat att bedömningar i samhällskunskap främst har fokuserat på att *minnas faktakunskaper* (Brookhart & Durkin, 2003; Curry & Smith, 2017; Jansson, 2011; Lindmark, 2013; Mattson, 1989; Odenstad, 2011). Jämfört med detta indikerar denna studie en ändrad bedömningspraktik bland samhällskunskapslärare, då lärarnas uppgifter i studien fokuserade mer på *förståelse* och resonerande förmågor som att *analysera* och *värdera*. Uppgifter som testar att *tillämpa* och *skapa* förekom också men i mindre grad. Lärarna testade också *procedurkunskaper*, främst källkritiska förmågor och samhällsvetenskaplig metod. Detta är mer i linje med samhällskunskap som ett diskussionsämne med kritiskt tänkande (Englund, 1986; Möllenborg, 2023; Torrez & Claunch-Lebsack, 2014; Wahlström m.fl., 2020).

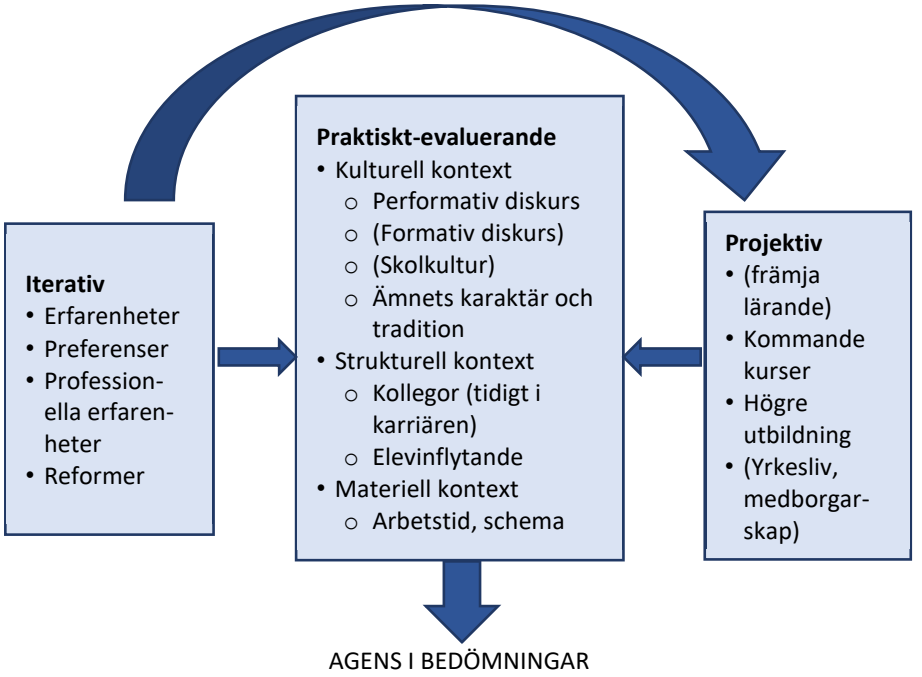
Resultaten visade också att lärarna varierar sina uppgifter. De använde såväl skriftliga som muntliga metoder, medan digitala former var ovanliga. Skriftliga uppgifter var dock mest använt av de flesta lärarna, även om det fanns några som föredrog mer muntliga metoder. Vanligast var skriftliga inlämningsuppgifter av olika slag, vilket alla lärarna använde. De flesta använde även skriftliga prov och på de högskoleförberedande programmen användes ofta rapporter. Av de muntliga metoderna var presentationer och seminarier vanligast. Andra metoder, till exempel poddar, rollspel och posters, användes mer sparsamt. Lärares metoder tycks således ha blivit mer varierade från att tidigare ha dominerats av skriftliga prov (Mattson, 1989; Odenstad, 2011).

Lärarna uttryckte inte någon tydlig relation mellan vilket innehåll och vilka förmågor som ska bedömas och val av metod. De hade svårt att uttrycka sina rationaler för sina beslut om bedömningar, vilket är i linje med tidigare studier (McMillan & Nash, 2000). Detta indikerar att

lärares beslut ibland kan vara intuitiva och oreflekterade (Edwards, 2020; Vanlommel m.fl., 2017), men det skulle också kunna vara ett bristande professionellt språk för bedömningar (Colnerud & Granström, 2015). Bedömning tycks vara en tyst kunskap som formas mindre av lärarutbildning och mer av lärares egna erfarenheter och preferenser (Amani m.fl., 2021; Campbell, 2013; Cooper m.fl., 2017). Samtidigt pekade lärarna på en del samband då de såg skriftliga prov som mer lämpliga för att testa ämnesområden med mycket fakta och för att testa att *minnas*. För att testa mer komplexa ämnesområden och resonerande förmågor, som att *förstå*, *analysera* och *värdera*, användes andra bedömningsformat, främst inlämningsuppgifter men också seminarier. När *procedurkunskaper* ska testas ansåg de det lämpligare med tillämpande uppgifter, till exempel att skriva rapporter. Även dessa val var dock främst grundade i lärarnas tidigare erfarenheter och egna preferenser (Edwards, 2020; McMillan, 2003; Öberg & Bäckström, 2021).

Lärarnas tidigare bakgrund i den iterativa (historiska) dimensionen (figur 7), har betydelse för deras agens i arbetet med bedömningar. Liksom i många andra länder (Amani m.fl., 2021; Campbell, 2013; Cooper m.fl., 2017) saknade de svenska lärarna i studien specifik utbildning i bedömning. Alla lärarna i studien gick sin lärarutbildning innan det blev obligatoriskt med kurser i bedömning. I stället formades deras bedöm-

Figur 7. *Läraragens i bedömningar. Från Priestley m.fl. (2015).*



ningspraktik utifrån deras preferenser och tidigare erfarenheter, både från egen skolgång och från lärararbetet, där de i början av sin yrkeskarriär ofta tagit intryck av kollegor. Avsaknaden av utbildning riskerar att lärarnas repertoar i bedömningar blir mer begränsad.

I den projektiva (framtidsinriktade) dimensionen tänkte lärarna långsiktigt om bedömningar när de planerar för hela kursen, främst på så sätt att de ville täcka in alla kunskapskrav, samt att varje kunskapskrav helst skulle testa minst två gånger för att ge eleverna möjlighet till att förbättra sina resultat. De ville också genom bedömningarna förbereda eleverna för kommande gymnasiekurser och för högre studier.

I den praktiskt-evaluerande (nutida) dimensionen är det flera kontextuella faktorer som har betydelse för läraragens. En stark betoning på betygsättning i styrdokumenten innebär en stark performativ diskurs (Strandler, 2017; Wahlström & Sundberg, 2015), en kulturell faktor, vilken har stor betydelse för lärarna då de måste ha underlag för betyg och täcka in alla kunskapskrav. Denna performativa diskurs tenderar att minska lärares agens (Bergh & Wahlström, 2018; Priestley m.fl., 2015). Det finns också en stark diskurs som förespråkar formativ bedömning i Sverige (Wahlström & Sundberg, 2015) men denna hade inte någon större betydelse för lärarnas val av innehåll, förmågor eller metoder. Lärarna talade inte så mycket om formativ bedömning när de talade om konstruktion av bedömningar. Det enda uttryckliga undantaget var att de flesta av lärarna inte ville använda poäng som återkoppling med motiveringen att det inte säger något om elevens kunskaper och inte kan användas formativt.

Tidigare studier beskriver motsättningar mellan en öppen och flexibel läroplan och en performativ diskurs (Priestley m.fl., 2015). När det gäller kunskapskraven uttryckte lärarna i studien denna motsättning men när det gäller det centrala innehållet hade lärarna en mer avslappnad inställning där de inte kände att de måste bedöma allt innehåll och de kände också en frihet i val av metoder. Detta bekräftar bilden av ämnet som öppet och flexibelt och som saknar ett tydligt innehåll (Hadley & Young, 2018; McMillan, 2001; Wahlström m.fl., 2020). Samtidigt pekade studien, i linje med annan forskning (Olsson, 2016; Öberg & Bäckström, 2021), på en tydlig ämneskanon, där också aktualisering ingår. Detta indikerar en stark kanonisk tradition som inramar ämnet, även om ämnesplanen är öppen och flexibel.

Det finns också strukturella faktorer som har betydelse för lärarnas agens. En sådan är elevernas behov och intressen, vilka kan påverka lärarnas beslut om framför allt metod. Lärarna kunde ibland låta eleverna, som grupp eller individuellt, välja format för bedömningar. Även gruppstorlek kan påverka val av metod (Duncan & Noonan, 2007; James & Pedder, 2006; McMillan, 2003) då vissa metoder, särskilt muntliga, en-

ligt några av lärarna kan vara svåra att genomföra i stora grupper. En annan strukturell faktor är relationen till kollegor, men lärarna uttryckte att förutom i början av karriären hade detta inte någon större betydelse för deras bedömningspraktik.

Till sist finns också de materiella villkoren, främst i form av schema, lärares arbetstid och elevernas arbetsbörda. Detta kunde påverka lärarnas beslut om antal uppgifter, innehåll och metoder men ibland också om enskilda bedömningsuppgifter. Några lärare uttryckte till exempel att de skulle vilja använda fler muntliga uppgifter men att de avstod på grund av tidsbrist. De materiella villkoren skapar spänningar till både krav i styrdokumenten och till lärares uppfattningar och preferenser, vilket begränsar lärares agens (Lumadi, 2013).

Sammanfattningsvis gav lärarna i samhällskunskap uttryck för att de kan uppnå en relativt hög grad av agens i arbetet med bedömningar. De har stor frihet att välja vilket innehåll som bedöms och med vilka metoder, en frihet som lärarna i studien nyttjade med en variation i metoder och olika fokus på innehåll. Om variationen mellan lärare blir för stor finns det dock en risk att eleverna blir bedömda på olika grunder. Eftersom bedömningar påverkar elevers lärstrategier (Brookhart & Durkin, 2003) riskerar de därmed att utveckla olika lärande, vilket kan innebära problem med likvärdig bedömning. Samtidigt kan denna frihet ge lärarna frirum att anpassa bedömningar utifrån elevernas förutsättningar och intressen men också till lärarnas egna intressen, vilket kan göra arbetet mer meningsfullt.

Lärare behöver balansera mellan en summativ och en formativ diskurs. De behöver också balansera detta och sina egna preferenser med administrativa krav i form av begränsad arbetstid och schema. Lärares agens skulle kunna vara större om lärare var mindre bundna av ämnestraditioner men också genom utbildning i bedömning, vilket skulle kunna ge en större bedömningsrepertoar. Detta skulle i sin tur ge större förutsättningar för högre validitet då lärare skulle vara bättre rustade för att kunna välja relevanta metoder beroende på vilket innehåll och vilka förmågor som ska bedömas.

Samhällskunskapslärarna i studien hade en mer varierad bedömningspraktik som i högre grad fokuserar resonerande förmågor, jämfört med äldre studier där skriftliga prov som testar att *minnas faktakunskaper* dominerar. Detta skulle innebära att samhällskunskapslärares bedömningar idag är mer valida i relation till styrdokumenten.

7 Diskussion

Arbete med bedömningar utgör en stor del av lärares arbete och det är en komplex och svår uppgift (Moss, 2013). Lärares bedömningar har också stor betydelse för eleverna. Lärarnas syften, vilket innehåll, vilka förmågor som testas och med vilka metoder påverkar elevernas lärande, men också deras motivation och självkänsla (Pekrun m.fl., 2023). Eftersom bedömningar används för urval har de också betydelse för elevernas framtid. Denna avhandlings delstudier har utforskat lärares bedömningspraktiker när det gäller deras beslut om syften med bedömningar, vilket innehåll och vilka förmågor de testar samt med vilka metoder. Lärares resonemang om relationerna mellan innehåll och förmågor som testas och val av metod har också undersökts. Vidare utforskades vilka faktorer som kan ha betydelse för lärares bedömningsbeslut samt lärares agens i relation till dessa faktorer.

När avhandlingen summeras framträder variation som centralt i samhällskunskapslärarnas bedömningspraktiker. Variation innebär både att de enskilda lärarna varierar sina bedömningsuppgifter och att det finns en variation mellan dem. Lärarna hade varierade och olika syften, de varierade vilket innehåll och vilka förmågor som testades och de varierade metoder. Lärarnas olika bedömningspraktiker har betydelse på individnivå (för den enskilde eleven), gruppnivå och samhällsnivå. Vilka kunskaper och förmågor som olika lärare bedömer, liksom deras val av metoder får också olika konsekvenser, då det på olika sätt kan påverka elevers motivation, självkänsla och lärstrategier. Variationen mellan lärarna har också betydelse för frågor om likvärdig bedömning och validitet. Dessa olika frågor diskuteras i avhandlingens avslutning.

Lärares bedömningspraktiker formas i en kontext där både historiska, nutida och framtida dimensioner har betydelse (Priestley m.fl., 2015). Ämnets karaktär är en av dessa kontextuella villkor. Inom dessa dimensioner, med olika rådande kontextuella villkor, uppnår lärare olika grader av agens för sina beslut om sina bedömningsuppgifter. Detta diskuteras också i avhandlingens avslutning.

I diskussionen skrivs lärarna i bestämd form för att markera att det är lärarna i studien som avses, inte lärare generellt. Däremot står det i obestämd form om det gäller lärare i allmänhet.

Varierade syften

Studien visade att lärarna hade olika och varierade syften med sina bedömningar: implementera policy, främja lärande, hjälpa elever att lyckas samt administrera bedömningar. Lärarna behövde balansera spänningar

mellan dessa olika syften. Samtidigt uttryckte lärarna en del syften som mer centrala än andra. Ett exempel är att lärarna betonar summativa syften med sina uppgifter, vilka de uttryckte att de oftast använde som underlag för betyg. Denna bild förstärks också av att lärarna ofta uttryckte att betygskriterierna har stor betydelse för deras beslut om bedömningar. Detta bekräftar andra studier som har visat att skolans styrdokument uttrycker en stark performativ diskurs (Strandler, 2017; Wahlström & Sundberg, 2015). Lärarnas betoning på syften att implementera policy innebär att deras bedömningspraktik tenderade att bli starkt orienterad mot summativ bedömning med fokus på betyg.

Samtidigt talade alla samhällskunskapslärarna i de senare delstudierna om formativa syften för att främja lärande, vilket är en skillnad mot den första delstudien, där lärarna inte talade om formativ bedömning. Dessa formativa syften påverkade dock inte deras konstruktion av bedömningsuppgifter i någon större utsträckning. Ett undantag var att flera av lärarna såg användningen av poäng som återkoppling i uppgifter irrelevant för bland annat formativa syften, då poäng inte säger något om kvaliteterna i elevernas prestationer och om hur de ska gå vidare, även om det fanns lärare som förespråkade användningen av poäng då det har viss legitimitet bland eleverna. Poängsättning har också lyfts fram av finska lärare som en olämplig praktik, eftersom det inte bidrar till lärande, även om en del finska lärare också trots allt använde det (Atjonen, 2014). Lärarna uttryckte att poäng fungerar dåligt i relation till det kriterierelaterade betygssystemet, då prestationer för olika kriterier klumpas i en poängsumma. Att en elev når en viss poänggräns behöver ju inte innebära att eleven visat kunskaper motsvarande alla kriterier som testas för ett visst betyg. Poäng är således problematiskt för att kunna göra valida tolkningar av elevprestationer både när det gäller att ge formativ återkoppling och för att sätta betyg.

Ett annat av lärarnas syften var att hjälpa eleverna att lyckas och lärarna uttryckte att variation, främst gällande metoder, är viktigt för att fler elever ska lyckas. De menade att variation kan hjälpa elever att lyckas, eftersom elever är olika och visar sina kunskaper på olika sätt. Variation av metoder betonas också i såväl litteratur om bedömning som av lärare i några andra studier (Atjonen, 2014; McMillan, 2018; McMillan & Nash, 2000). Forskning har också visat att bedömningsuppgifternas utformning kan gynna eller missgynna olika elever och det kan också påverka elevernas känslor och självtillit, antingen positivt eller negativt (Löfgren m.fl., 2019; Pekrun m.fl., 2023). Dessa konsekvenser är något som lärare också behöver beakta i sina beslut om bedömningar, vilket diskuteras mer i avsnittet om konsekvenser.

Ett sätt att hjälpa eleverna att lyckas är, enligt lärarna, att eleverna får påverka både när och hur bedömningar ska genomföras. Elevinflyt-

ande är också framskrivet i läroplanen (Skolverket, 2022b). Samtidigt är elevinflytande på bedömningar inte oproblematiskt. I studierna uttryckte en del lärare att de använde en del uppgifter för att eleverna önskar dem, även om lärarna egentligen inte såg uppgifterna som de mest ändamålsenliga. Ett exempel är att lärarna ibland använde skriftliga prov i högre utsträckning med motiveringen att dessa har legitimitet hos eleverna och för att de är vana vid dem. Här finns således en spänning mellan styrdokumentens krav på elevinflytande och ändamålsenliga underlag för betygsättning, men också gentemot lärares professionella omdöme där lärarna behöver balansera sin önskan att hjälpa eleverna och att följa styrdokumenten.

Slutsatsen är att även om samhällskunskapslärarna hade varierade syften var deras bedömningspraktik starkt summativt orienterad. Formativa syften påverkade i mindre grad deras utformning av bedömningsuppgifter. En variation av metoder innebär också att olika elevers behov kan tillgodoses och fler kan få möjlighet att lyckas. Samtidigt finns det en risk att elever utvecklar olika kunskaper och förmågor och att de bedöms på olika grunder, vilket kan påverka likvärdighet och rättvisa. Det finns spänningar mellan lärarnas olika syften, vilka lärarna i sin professionsutövning behöver balansera. Olika syften komplicerar också frågan om validitet i deras bedömningar, eftersom lärares uppgifter behöver vara relevanta i relation till deras syften (Kane & Wools, 2020; Newton & Shaw, 2016). Olika kontextuella villkor ger dock lärare olika agens att hantera dessa spänningar. Detta diskuteras mer nedan.

Varierade bedömningspraktiker

Samhällskunskapslärarna i studien hade även en varierad bedömningspraktik, både vad gäller vilket innehåll som testas, vilka förmågor, och med vilka metoder. Samtidigt varierar praktikerna mellan lärarna. Styrdokumenten föreskriver inte hur lärarna ska testa elevers kunskaper och förmågor, annat än att eleverna i ämnet samhällskunskap måste visa sina kunskaper både muntligt och skriftligt (Skolverket, 2022c). Det ger lärarna stort frirum och agens att välja metoder för bedömning. Detta gäller också vilket innehåll som ska bedömas. Visserligen stipuleras ett centralt innehåll i samhällskunskap, vilket i de studerade kurserna är brett, men det uttrycks inte att allt innehåll måste testas. Det är också så lärarna i studien uppfattade det centrala innehållet, då de uttryckte att allt innehåll inte behöver testas. Däremot uttryckte de att de måste ha underlag för alla betygskriterier och de ville helst testa varje betygskriterium minst två gånger. Detta är också något som styrdokumenten uttrycker (Skolverket, 2018, 2022a), vilket bidrar till betoningen på en summativ bedömningspraktik (Wahlström & Sundberg, 2015).

I enlighet med denna frihet utformade också lärarna olika bedömningspraktiker, där de hade olika preferenser för olika metoder. En del föredrog skriftliga prov, andra skriftliga inlämningsuppgifter och ytterligare andra muntliga uppgifter. Alla varierade dock sina uppgifter och använde både skriftliga och muntliga former men i olika grad. De lade olika tonvikt vid vilket centralt innehåll som testas, även om vissa områden var vanligare att samhällskunskapslärarna bedömde. (Detta diskuteras mer i avsnittet om ämnets karaktär.) Alla lärarna testade också olika förmågor, både att *minnas*, *förstå* och resonerande förmågor som att *tillämpa*, *analysera* och *värdera*, men de gjorde det i olika grad. Främst var det en betoning på resonerande förmågor, även om några av lärarna lade större tonvikt vid minneskunskaper.

Lärarnas variation av metoder är i likhet med lärare i nederländska studier (Kippers m.fl., 2018) men skiljer sig från amerikanska studier där lärarna hade en mindre variation (Brookhart, 2013; Brookhart & Durkin, 2003; Curry & Smith, 2017). Bedömning har dock även i Sverige tidigare varit mindre varierad och med en tonvikt på skriftliga prov (Mattson, 1989; Svingby, 1998). Den performativa diskursen skulle kunna ha bidragit till denna mindre varierade bedömningspraktik, men bedömningspraktiker med fokus på faktaorienterade skriftliga prov fanns redan sedan tidigare (Lundahl, 2020). Det skulle också kunna bero på olika bedömningskulturer, där Sverige har påverkats mer av en anglosaxisk psykometrisk tradition, medan till exempel Norge har följt en mer kontinental tradition (Tveit, 2018). Det är möjligt att även Nederländerna har påverkats av den kontinentala traditionen. Enligt Tveit underlättade den svenska bedömningskulturen att en performativ diskurs snabbt fick genomslag i Sverige. Ytterligare en förklaring är att lärare har uttryckt att de föredrar lätträttade prov för att summera kunskaper framför uppgifter som kräver en subjektiv bedömning (Törnvall, 2001). Även detta är ett uttryck för den dominerande summativa praktiken och dess konsekvenser. Samhällskunskapslärarna i denna studie har dock, som sagt, mer varierade bedömningspraktiker, vilket skiljer sig från bilden i andra ämnen i Sverige. Detta diskuteras i nästa avsnitt om ämnets karaktär.

Slutsatsen är att styrdokumenten både begränsar och möjliggör läraragens i arbetet med bedömningar. Lärarna uttrycker ett större frirum och agens i relation till val av innehåll och metoder, medan de uttrycker en snävare agens när det gäller beslut om förmågor utifrån kraven på att ha underlag för alla betygskriterier. Att lärarna testar olika innehåll kan leda till att elever lär sig olika innehåll. Det har också betydelse för validitet, vilket diskuteras nedan.

Ämnets karaktär möjliggör variation

Delstudierna har visat att samhällskunskapslärarna, förutom att de varierade sina bedömningsmetoder, också betonade resonerande förmågor i sina bedömningar. Detta innebär att deras bedömningar är i linje med ämnets reflekterande karaktär (Arevik & Hartzell, 2015; Barton & Avery, 2016; Lindqvist & Hyltegren, 1995; Svingby, 1989; Torrez & Claunch-Lebsack, 2014). Här skiljer sig samhällskunskapslärarna från lärarna i andra studier gällande naturvetenskapliga ämnen, teknik och historia, vilka i högre grad använde skriftliga prov som testade faktakunskaper (Gomez, 2017; Hartell & Strimel, 2019; Kakoulidou, 2021; Rosenlund, 2016). Naturvetenskapliga ämnen betraktas som mer slutna och hierarkiska ämnen med stark inramning (Bernstein, 2003; Biberman-Shalev m.fl., 2011; Linde, 2021). Det är möjligt att även historia har en sådan karaktär. I vart fall verkar det finnas en stark ämnestradition (Rosenlund, 2016) som konserveras genom att den överförs mellan olika generationer av lärare (Elwood & Murphy, 2015). Samhällskunskap beskrivs däremot som ett öppet och flexibelt ämne med svag inramning (Linde, 2021; McMillan, 2001), vilket skulle kunna förklara skillnaden gentemot de andra ämnena. Det vore dock intressant med fler studier om detta.

Lärarna nyttjade främst ämnets flexibilitet och öppenhet när det gäller ämnets centrala innehåll och val av metoder. Däremot hade betygskriterierna stor betydelse för hur lärarna utformade sina uppgifter. Styrdokumenten ingår i läraragens som en del i den iterativa dimensionen där reformer ses som en del av historien. Samtidigt utgör betygsreformen 2011, enligt Wahlström och Sundberg (2015), en stark performativ diskurs, vilken är en kulturell faktor i den praktiskt-evaluerande dimensionen. Betygskriterierna hade stor betydelse för såväl lärarnas syften med bedömningar som deras beslut om bedömningar, framför allt om vilka kunskapsformer som ska bedömas, men i viss mån också med vilka metoder. Till exempel uttryckte lärarna att de prioriterade skriftliga uppgifter framför muntliga då de är lättare att använda för summativa syften för att ”ha ryggen fri”, vilket är en tydlig påverkan av den performativa diskursen. Lärarna upplevde också att betygskriterierna i samhällskunskap är problematiska. De var särskilt frustrerade över de innehållsspecifika kriterierna om mänskliga rättigheter och privatekonomi, vilka de upplevde att de måste bedöma på något sätt. De menade också att dessa kriterier riskerar att bli avgörande för elevernas betyg.[7]

[7] Detta utifrån besiktningsmodellen som gällde mellan 2011 och 2022 då alla kunskapskrav för ett visst betyg (E, C eller A) behövde vara uppfyllda för att eleven skulle få det betyget. Efter 2022 gäller kompensationsmodellen där brister i några betygskriterier kan vägas upp med styrkor i andra och lärarna gör en sammantagen bedömning (se avsnittet om betygshistorik).

Samtidigt som lärarna var flexibla i vilket innehåll som bedöms har studier visat att det finns en stark kanonisk tradition (Öberg & Bäckström, 2021) som är en annan kulturell diskurs som influerar lärarnas bedömningsuppgifter. Det var vanligare att vissa arbetsområden bedömdes, både utifrån vad lärarna uttryckte och i de insamlade uppgifterna, nämligen statsvetenskap (om politiska system, främst om hur Sverige styrs) och nationalekonomi. Detta är också i linje med tidigare studier (Karlefjärd, 2011; Karlsson, 2011; Odenstad, 2011; Olsson, 2016; Öberg & Bäckström, 2021). Styrdokumentens betydelse slår också igenom i val av innehåll då en skillnad mot ovannämnda studier är att lärarna, på grund av de ämnesspecifika kunskapskraven, i större utsträckning bedömde mänskliga rättigheter. Källkritik är också tydligare framskrivet i ämnet, samtidigt som det sågs som en del av ämnets karaktär. Lärarna uttryckte också att den är viktig som förberedelse för kommande studier och för medborgarfostran (en projektiv dimension i läraragens). Aktualisering är ett annat karaktärsdrag som lärarna uttryckte och därför relaterade de ofta uppgifter till dagsaktuella händelser i sina uppgifter.

Wahlström m.fl. (2020) menar att kursplanen i samhällskunskap betonar generella förmågor mer än ämnesspecifika förmågor. Till exempel menar de att analysera uttrycks som en generell förmåga. Liksom lärarna i Möllenbergs (2023) studie hanterade dock lärarna i denna studie inte dessa förmågor som generella. När de gav eleverna i uppgift att analysera så skedde detta alltid i ett ämnesspecifikt sammanhang och de kunde även ange en ämnesspecifik analysmodell som skulle användas i analysen. Det enda undantaget var en lärare som menade att han kunde använda elevers prestationer i andra ämnen, till exempel naturkunskap, för att bedöma elevens analysförmåga i samhällskunskap. Denna lärare var dock också den enda som inte hade en ämnesutbildning i samhällskunskap, vilket skulle kunna förklara denna skiljaktiga praktik. Således var de studerade samhällskunskapslärarnas praktik förankrad i ämnet, även om styrdokumenten skulle vara generellt formulerade.

Slutsatsen är att ämnets öppenhet och flexibilitet ger samhällskunskapslärarna stort frirum att välja innehåll och metoder, vilket möjliggör en hög grad av agens. Ämnets karaktär påverkar lärarnas bedömningspraktiker avseende innehåll, i form av aktualisering, vilka förmågor som bedöms genom betoningen på resonerande uppgifter och källkritik samt en del beslut om metoder. Samtidigt påverkar en ämnestradition vilket innehåll lärarna bedömer. Att ämnets resonerande karaktär avspeglas i lärarnas bedömningspraktiker är också positivt för validiteten, vilket diskuteras mer i nästa avsnitt.

Variation och validitet

Samhällskunskapslärarnas variation av syften, testat innehåll och förmågor samt metoder påverkar också validiteten i deras bedömningar. För att lärare ska kunna dra valida slutsatser om elevers kunskapsutveckling behöver bedömningsuppgifter vara relevanta i relation till de syften som lärarna har med respektive bedömning (Kane & Wools, 2020; Messick, 1989). Validitet har betydelse för såväl summativ som formativ bedömning då lärarna behöver valida underlag för att kunna ge relevant återkoppling eller fatta relevanta beslut om fortsatt undervisning, liksom för betygsättning (Dolin m.fl., 2018; Gu, 2021; Wellberg, 2023). För valida bedömningar behöver lärare välja lämpliga metoder beroende på vilket innehåll och vilka förmågor som ska mätas (Kane & Wools, 2020; McMillan, 2018; Messick, 1989).

Lärarna i studien talade inte uttryckligen om validitet. De hade ofta först svårt att peka på samband mellan lämpliga metoder i relation till vilket innehåll eller förmågor som ska testas, vilket är i likhet med de amerikanska lärarna i McMillans och Nashs (2000) studie, vilka också hade svårt att förklara grunderna för sina val. Dock framkom några samband som att prov fungerar bättre för att mäta minneskunskaper, medan inlämningsuppgifter och andra format är bättre för mer resonerande frågor. Några av lärarna använde dock prov med bara essäfrågor som testade fler förmågor än att *minnas* och det var få som hade rena faktaprov. I relation till samhällskunskapsämnets karaktär fanns en relativt god överensstämmelse mellan lärarnas uppgifter och de förmågor som skulle testas. Samtidigt varierade bedömningspraktikerna mellan lärarna, vilket innebär att graden av validitet varierar.

Lärarnas fokus på att ha underlag som testar det som står i betygskriterierna ligger väl i linje med validitet i relation när det gäller betygsättning. Dock visar delstudie I en diskrepans mellan lärarnas uppfattningar av betygen och deras prov. Deras förståelse av de då gällande betygskriterierna (2009) var i linje med styrdokumenten, medan proven inte helt överensstämde med de betygskriterier som de avsåg att testa. Framför allt överbetonades faktakunskaper. Det fanns dock en variation mellan lärarna i studien där några prov har en större diskrepans gentemot betygskriterierna än andras. En del av lärarnas prov var därmed inte helt representativa för de förmågor som de avsåg att mäta. Samtidigt drog dessa lärare slutsatser om betyg som de egentligen inte hade underlag för, vilket innebär problem med validiteten (Dolin m.fl., 2018; Kane & Wools, 2020). Det innebär också en diskrepans mellan lärarnas förståelse av betygskriterierna och deras praktik, något som också visats i andra studier (Black m.fl., 2010). Delstudie I grundades dock endast på lärarnas prov och delstudie III ger en mer nyanserad bild än delstudie I,

liksom några andra studier där lärarna främst använde skriftliga prov som testade faktakunskaper (Brookhart & Durkin, 2003; Curry & Smith, 2017; Lindmark, 2013; Mattson, 1989; Odenstad, 2011). I delstudie III använde lärarna en variation av uppgifter och lärarna testade i högre grad resonerande förmågor som att *förstå*, *analysera* och *värdera*, medan att *minnas* är mindre betonat. Detta möjliggör en mer valid bedömningspraktik i relation till styrdokumenten och kunskapskraven, och därmed också till lärarnas summativa syften, eftersom bedömningsuppgifterna ofta används som underlag för betyg.

En intressant iakttagelse är att lärarna dock var mer angelägna om att bedöma elevernas förmågor utifrån vad som står i betygskriterierna, än det centrala innehållet som de inte tyckte var lika viktigt att täcka in. Det indikerar att för lärarna gäller validiteten främst förmågorna, medan innehållet är mindre viktigt. Detta innebär en risk för underrepresentation av det innehåll som ska testas.

Lärarna uttryckte att de har formativa syften med sina bedömningar, men som nämnts ovan påverkade det inte hur de utformade sina bedömningar. I stället var det snarare utifrån summativa syften som de utformade sina bedömningar, samtidigt som de ibland använde uppgifterna formativt. Detta kan påverka validiteten negativt i relation till formativa syften då en kombination av summativ och formativ återkoppling riskerar att reducera effekten på lärandet (Lauvås m.fl., 2019; Shute, 2008). Elever tenderar att inte ta till sig formativ återkoppling om den ges tillsammans med summativa omdömes som betyg.

Samtidigt behöver lärarna även beakta sina andra syften som att hjälpa elever att lyckas och att hantera administrationen av bedömningarna. Lärarna behöver således kompromissa mellan olika syften (Kane, 2016; Newton, 2017), vilket kan innebära att om huvudsyftet snarare är att främja lärande eller att hjälpa eleverna att lyckas, kan de behöva göra avkall på validiteten i relation till styrdokumenten. Likaså kan de behöva balansera mellan administrativa och andra syften då både lärares och elevers begränsade tid kan kräva kompromisser. Det är i denna process som lärare måste prioritera mellan sina syften, vilket avgör vad som blir viktigast i relation till validitet.

Lärarna hade som mål att testa de olika kunskapskraven flera gånger, vilket också är positivt för validiteten, eftersom flera underlag är önskvärt för en mer valid tolkning och användning. Variation minskar risken för irrelevans, att bedömningen inte testar det som avses eller att andra aspekter påverkar bedömningen (Messick, 1989). Variationen behövs också för konvergent evidens, det vill säga att samma förmåga testas med olika metoder och får samma resultat. Samtidigt begränsas möjligheterna att ha flera bedömningar av lärarnas kontextuella villkor såsom arbetstid och schema.

Slutsatsen är att validitet inte tycks vara något lärarna har reflekterat över. Detta innebär en risk att lärarna inte alltid utformar sina uppgifter på ett ändamålsenligt sätt så att de mäter det som de avser att mäta. Utbildning för såväl lärarstudenter som yrkesverksamma lärare skulle kunna öka lärares förståelse för validitet liksom deras kompetens i att konstruera valida bedömningar utifrån olika syften, det vill säga deras konceptuella och praxeologiska bedömarliteracitet (Pastore & Andrade, 2019). Samtidigt tenderade samhällskunskapslärarnas bedömningar att vara mer valida för summativa syften, än för formativa syften.

Konsekvenser av varierade bedömningspraktiker

Ovan konstaterades att lärarna i studien hade en varierad bedömningspraktik men också att den var olika mellan dem. Tidigare forskning har pekat på att lärares bedömningspraktiker påverkar eleverna på olika sätt. Det påverkar vad eleverna uppfattar som viktigt att lära, deras lärstrategier, motivation, självförtroende liksom känslor för bedömningar (Brookhart & Durkin, 2003; Löfgren m.fl., 2021; Pekrun m.fl., 2023). Olika bedömningspraktiker har också betydelse för likvärdighet (Korp, 2011) och för konsekvensvaliditet (Messick, 1989). I beaktandet av konsekvenser behöver lärarna ta hänsyn till både principen om *avoid score pollution* och *do no harm* (Green m.fl., 2007).

Ovan konstaterades att samhällskunskapslärarna bedömde olika delar av kursens centrala innehåll och att de lade olika tonvikt på betygskriterierna, där några mer betonade minneskunskaper och andra mer resonerande förmågor. Vad som bedöms och hur har betydelse för elevernas lärstrategier och vad de lär sig (Brookhart & Durkin, 2003; Moss, 2013) Detta innebär att lärarnas olika fokus i sina bedömningspraktiker kan leda till att eleverna lär sig olika saker och utvecklar olika förmågor, där en del elever främst fokuserar på minneskunskaper, medan andra fokuserar på resonerande förmågor. Det finns också en risk att eleverna får betyg på olika grunder, vilket är problematiskt för likvärdigheten. Samarbete och samtal mellan lärare skulle kunna förbättra likvärdigheten, även om effekten av sådana samtal inte är helt entydig (Jönsson & Thornberg, 2014). Lärarna uttryckte dock att det sällan förekom samarbete om bedömningar.

Lärarna betonade summativa syften. Detta innebär en risk att både lärare och elever utvecklar en instrumentell syn på skolan med fokus på betyg snarare än lärande (Grettve m.fl., 2014; Löfgren m.fl., 2021; Olovsson, 2015). Betyg, vilket lärarna tenderade att ge även på enskilda uppgifter, riskerar att minska elevernas inre motivation och även skada lärandet. De påverkar också elevernas självförtroende och känslor, där elever som lyckas stärks, medan elever som misslyckas riskerar att

hamna i en ond cirkel av sämre självförtroende och negativa känslor inför bedömningar.

Lärarnas varierade bedömningspraktiker är positivt i relation till konsekvenser. Variationen gör det möjligt att möta olika elevers behov och intressen. Det minskar risken att en del elever upplever all bedömning som negativ eller meningslös och ökar sannolikheten att de känner motivation och självförtroende i en del bedömningar. Förhoppningsvis kan det också spilla över på andra uppgifter. Lärarnas betoning på summativ bedömning kan ses som positivt utifrån principen *avoid score pollution*, men samtidigt är den problematisk utifrån principen *do no harm*. Lärare tvingas balansera dessa principer vid beaktandet av konsekvenser av bedömningar.

Sammanfattningsvis är det viktigt med en medvetenhet om bedömningspraktikers konsekvenser för eleverna. Det finns en del utmaningar för lärare i att hantera negativa konsekvenser av bedömningar och i stället stärka positiva effekter för att öka elevers motivation och självförtroende samt främja lärande av olika förmågor.

Samhällskunskapslärarnas agens

Samhällskunskapslärarna uttryckte, liksom Moss (2013), att bedömning är komplext och svårt. Lärarna har många beslut att fatta när de konstruerar sina bedömningsuppgifter. De hade många olika syften, både interna och externa, som de behöver balansera. Samtidigt formades deras bedömningspraktiker av deras bakgrund, ambitioner och rådande kontextuella villkor. Vad som är att betrakta som en ändamålsenlig och valid bedömning blir i denna kontext också komplext då en bedömning kan vara mer eller mindre valid i relation till olika syften. Dessutom behöver lärarna ta hänsyn till bedömningarnas konsekvenser.

Det finns en del faktorer som inramar lärarnas arbete med bedömningar. Det som starkast tycks begränsa deras frirum och agens är en performativ diskurs uttryckt främst i kursplanens betygskriterier, vilka lärarna tog stor hänsyn till. Vidare har materiella villkor betydelse då lärarna behöver hantera sin arbetstid och eleverna måste hinna med att lösa uppgifterna. I övrigt tycks lärarnas agens vara stor i relation till olika kontextuella villkor. Det är snarare inre faktorer som påverkar lärarna mest, i form av deras värderingar och preferenser. Detta har formats i den iterativa dimensionen genom deras erfarenheter både som elev och lärare, men också socialisering genom kollegor under de första åren, medan utbildning hittills spelat mindre roll. Det skulle kunna vara så att lärare som har påbörjat sin utbildning efter 2011, när kurser i bedömning blev obligatoriska, i högre grad har påverkats av sin utbildning, då andra studier visar positiva effekter av utbildning i bedömining (Koloi-Keaikit-

se, 2016; Zhang & Burry-Stock, 2003). Huruvida svenska lärare som gått lärarutbildning efter 2011 påverkats i sin bedömningspraktik återstår dock att undersöka.

Ämnets öppna och flexibla karaktär ger samhällskunskapslärarna stort frirum att forma sina bedömningspraktiker i relation till deras olika syften med bedömningarna. Elevernas bakgrunder och förutsättningar utgör en del av den strukturella kontexten för lärarnas agens. Utifrån syftet att hjälpa elever att lyckas kan lärarna anpassa bedömningarna efter elevernas behov och intressen, vilket de uttryckte att de på olika sätt gör, bland annat genom den konstaterade variationen i lärarnas bedömningspraktiker. Det ger också möjlighet för lärarna att forma sin praktik utifrån sina egna preferenser. Samtidigt påverkas lärarna mer eller mindre av ämnestraditionen vad gäller vilket innehåll som bedöms, medan den tidigare traditionen med skriftliga prov som testar minneskunskaper inte längre är lika självklar. I stället varierade lärarna metoder och testade mer resonerande förmågor. I relation till validitet innebär variationen också att samhällskunskapslärarnas bedömningar ger underlag för valida tolkningar och användningar.

Detta kan dock problematiseras och studien indikerar också att den starka performativa diskursen har störst betydelse för lärarnas bedömningspraktik. I avhandlingen framträder att mycket av samhällskunskapslärarnas arbete var fokuserat på att ha underlag för att kunna sätta betyg. Även om inte lärarna explicit talade om validitet, gjorde de det implicit genom att de var angelägna om att konstruera uppgifter som kunde testa alla betygskriterierna. Detta kan ses som positivt utifrån ett myndighetsperspektiv där lärarna som goda byråkrater professionellt och utan egna intressen implementerar det som står i styrdokumenten. Dock är knappast betygsättningen skolans primära uppdrag och syfte, utan snarare elevernas lärande och demokratiska fostran. Bedömning och betygsättning är ju inte ett mål i sig, utan snarare ett medel. I praktiken tycks det dock vara tvärtom, liksom uttrycket att bedömning är ”the tail that wags the curriculum dog” (Hargreaves, 1989), med vilket menas att bedömning, som egentligen borde vara en bisak, i praktiken blir det som styr undervisning och lärande, snarare än läroplaner. Eftersom bedömning har visats ha stor betydelse för vad eleverna lär sig (Brookhart & Durkin, 2003; Moss, 2013) påverkar således lärares bedömningspraktik både den enskilda eleven och indirekt samhället genom de kunskaper som eleverna utvecklar i skolan. Elever i Sverige riskerar att utveckla en instrumentell syn på bedömning där fokus är mer på att bocka av uppgifter än på lärande (Grettve m.fl., 2014; Löfgren m.fl., 2021; Olovsson, 2015).

Detta fokus på att testa betygskriterierna skulle inte behöva vara ett problem om lärare utformar sina uppgifter så att de testar de kunskaper

är uttryckta i styrdokumenten. Om eleverna anpassar sina lärstrategier efter sådana uppgifter skulle det leda till att de fokuserar på de kunskaper och förmågor som de förväntas lära sig. I praktiken visar avhandlingen att det är många aspekter som lärarna behöver väga in, även administrativa och elevernas önskemål, vilket komplicerar det hela. Likaså påverkar andra kontextuella villkor, som ämnestraditioner och lärarnas egna tidigare erfarenheter av bedömning. Dessa traditioner verkar vara ganska starka och praktiker som användning av poäng och fokus på minneskunskaper, vilka var rationella i det normrelaterade betygssystemet, har visat sig överleva länge, trots att de är svåra att förena med såväl ett kriterierelaterat betygssystem som med formativ bedömning. I samhällskunskap indikerar dock studien att lärares bedömningspraktiker med tiden kommit att bli alltmer i överensstämmelse med styrdokumentens kunskapssyn både generellt och specifikt i ämnet. Avsaknaden av utbildning i bedömning har troligen bidragit till konserveringen av tidigare bedömningspraktiker och den nuvarande lärarutbildningen, där undervisning om bedömning ingår, kan bidra till en utveckling av lärares bedömningspraktiker. Kompetensutveckling om bedömning skulle kunna bidra till att även höja redan yrkesverksamma lärares bedömarkompetens.

Studien indikerar att samhällskunskapslärare på gymnasiet har möjlighet att uppnå en hög grad av agens utifrån den relativt öppna läroplanen och kursplanen. Samtidigt tyder studien på att lärarna inte till fullo nyttjar detta frirum, utan begränsas av traditionerna som nämnts ovan. Inom ramfaktorteori beskrivs det upplevda frirummet, ett perspektiv som också är fruktbart att föra in i teorin om läraragens. Det innebär att de reella kontextuella villkoren kan medge en större läraragens än vad lärarna upplever. En utmaning för lärare kan således vara att upptäcka detta frirum och de möjligheter som finns. Lärares repertoar har också betydelse för deras agens (Priestley m.fl., 2015), det vill säga deras kunskaper om ämnet och om bedömning. Lärares bristande utbildning i bedömning riskerar att de har en snävare repertoar inom bedömning, vilket skulle kunna begränsa deras agens. Utbildning i bedömning kan således också bidra till att öka lärares bedömningsrepertoar och därmed deras agens när det gäller bedömningar.

Den sammantagna bilden är att lärarna anstränger sig för att konstruera uppgifter som ska testa de förmågor som är uttryckta i framför allt betygskriterierna, medan det centrala innehållet inte betonas på samma sätt. I delstudie I drogs slutsatsen att lärare kan och vill göra valida bedömningar. Denna slutsats bör dock revideras till att lärare vill och *skulle kunna* göra valida bedömningar under gynnsamma kontextuella villkor samt med ökad kunskap om konstruktion av ändamålsenliga bedömningar. Det behövs dock också en medvetenhet om att bedömningar

inte kan vara valida för alla olika syften samtidigt, utan lärarna behöver prioritera vilka syften som är viktigast (Kane & Wools, 2020). De behöver också ta hänsyn till bedömningarnas konsekvenser genom att minimera negativa konsekvenser och förstärka positiva. Detta kan dock komma i konflikt med validiteten i relation till en del syften, till exempel betygsättning, medan andra syften som att främja lärande och hjälpa elever att lyckas leder till mer positiva konsekvenser. Detta kan dock också problematiseras då en för stark tonvikt på dessa sistnämnda syften kan leda till problem med likvärdig bedömning. I slutändan måste lärarna balansera alla dessa olika krav i sin bedömningspraktik. Detta bekräftar bilden av att bedömning är svårt och komplext, men också ett viktigt arbete, där lärarna har möjlighet till stor agens och professionalitet.

Studiens begränsningar och fortsatt forskning

Denna studie bidrar till kunskaper om svenska samhällskunskapslärares bedömningspraktiker på gymnasiet. Kunskapen om hur lärare faktiskt gör och tänker om bedömningar är dock fortfarande mycket begränsad, både i Sverige och internationellt (McMillan, 2013; Torrez & Claunch-Lebsack, 2014) och det behövs mer forskning.

I avhandlingen används kvalitativa metoder i form av semistrukturerade intervjuer och analys av insamlade bedömningsuppgifter från ett relativt litet antal lärare, sex i den första datainsamlingen, tretton i den andra. Det innebär att det inte går att generalisera studiens resultat kvantitativt. Däremot går det att använda resultaten för generaliseringar av kvalitativa resultat (Larsson, 2010). Urvalet av informanter har gjorts strategiskt utifrån att få en så stor spridning som möjligt i lärarnas bakgrund och sammanhang, liksom utifrån deras inskickade material (Trost & Hultåker, 2016). Ett sådant strategiskt urval möjliggör att studierna fångar en stor bredd av den variation som kan finnas i samhällskunskapslärares bedömningspraktik och deras uppfattningar om denna (Larsson, 2010). I analyserna av både intervjuer och lärarnas uppgifter uppstod också en mättnad i koderna efter att drygt hälften av materialet analyserats. Detta stärker tillförlitligheten i delstudiernas olika kvalitativa teman, även om det skulle kunna finnas fler teman. Resultaten bör därför kunna appliceras i liknande ämnen och kontexter, en form av kvalitativ generalisering som Larsson (2010) benämner som en generalisering utifrån kontextlikhet och igenkänning av mönster. Främst bör resultaten kunna användas för att analysera eller förstå andra svenska samhällskunskapslärares bedömningspraktiker, men de kan också tänkas användas i andra länder, liksom i andra ämnen, där det kan finnas liknande mönster. Det är därför önskvärt med fler liknade studier, i såväl samhällskunskap, i andra ämnen, som på andra stadier.

Studien bygger på en mixad metod, där både insamlade artefakter i form av lärares bedömningsuppgifter och intervjuer används. Detta är också en styrka och möjliggör en triangulering mellan dessa olika underlag (Larsson, 2005). De insamlade uppgifterna är endast lärarnas skriftliga uppgiftsinstruktioner. Det hade också varit intressant att studera lärarnas muntliga uppgifter, men detta hade varit tidskrävande och svårt att praktiskt arrangera och valdes därför bort. Dessa uppgifter kommer dock till viss del med genom intervjuerna med lärarna men det är främst vilken typ av muntlig uppgift som tas upp, medan vilken typ av förmåga eller vilket innehåll som dessa muntliga uppgifter behandlar sällan finns med i utsagorna. Internationellt har många studier gjorts med kvantitativa metoder, främst enkätstudier, vilket inte tycks ha gjorts i Sverige sedan Mattsons stora studie (1989). Det kunde därför vara intressant med liknande kvantitativa studier för att få en större bredd och representativitet i bilden av lärares bedömningspraktiker. Samtidigt innebär sådana studier, liksom intervjuer, att studierna grundar sig på lärares egna självskattningar. Det vore således också intressant med fler studier som liksom denna avhandling använder artefakter i form av lärares egna uppgifter för att studera lärares bedömningspraktiker.

Ämnet samhällskunskap har tidigare konstaterats ha en svag inramning (Bernstein, 2003; Biberman-Shalev m.fl., 2011; Öberg & Bäckström, 2021). Även om denna avhandling pekar på att denna bild kan nyanseras, finns det fortfarande en del som pekar i den riktningen. Även detta kan studeras vidare och att analysera lärares bedömningsuppgifter är en viktig del i sådana studier då dessa signalerar vad som är viktig kunskap i ämnet samhällskunskap. Sådana studier bör inte begränsas till faktainnehållet, utan även analysera vilka typer av förmågor som lärare testar i sina uppgifter. Andra studier indikerar att undervisning och bedömning i samhällskunskap skiljer sig beroende på om det sker på yrkesprogram eller högskoleförberedande program (Möllenborg, 2023; Odenstad, 2011). Det vore intressant med fler studier om detta. Liknande studier i andra ämnen vore också önskvärt, både för att analysera vad som ses som centralt i ämnena, och för att kunna göra komparativa studier. För jämförbarhet skulle Blooms reviderade taxonomi kunna vara ett användbart verktyg (Näsström, 2008). Komparativa studier skulle eventuellt kunna identifiera olika karaktärer i lärares bedömningspraktiker mellan ämnen, både vad gäller vilken typ av uppgifter som används och vilka förmågor som betonas.

I denna studie förekommer digitala former av bedömning i liten utsträckning. Flera av lärarna nämnde under intervjuerna att det var på gång att införas digitala examinationsverktyg på skolorna. Under hösten 2022 och våren 2023 har det också blossat upp en debatt utifrån utvecklingen av AI (artificiell intelligens), där nya verktyg kan producera

texter, vilket underlättar och ökar risken för fusk vid inlämningsuppgifter. Detta har föranlett att Skolverket våren 2023 gav lärare rådet att inte använda uppgifter som eleverna jobbar med hemma som underlag för betyg (Skolverket, 2023). Det vore intressant att undersöka om och hur dessa digitala verktyg och AI påverkar lärares bedömningspraktik. Det skulle till exempel kunna tänkas att inlämningsuppgifter minskar i användning på grund av AI-verktyg, eller att de görs på annat sätt, exempelvis i digitala examinationsverktyg. De digitala examinationsprogrammen skulle också kunna påverka formen för lärares uppgifter och till exempel skulle självrättande frågor kunna bli vanligare.

Implikationer

Bedömning är en central och viktig uppgift för lärare och den har stor betydelse för de enskilda eleverna men också för samhället. Kunskapen om lärares bedömningspraktiker, både generellt och specifikt i samhällskunskap, är dock begränsad (McMillan, 2013; Torrez & Claunch-Lebsack, 2014) och denna avhandling bidrar med kunskap om detta.

Bedömning har länge dominerats av användning av prov som främst testar faktakunskaper, även efter skolreformerna 1994 med en ny kunskapssyn som betonade andra kunskapsformer. Detta tycks ha haft svårt att slå igenom i lärares bedömningspraktiker, men denna avhandling indikerar att det har skett en förändring åtminstone i ämnet samhällskunskap. En sådan förändring har dock tagit lång tid, vilket är viktigt att ta i beaktande vid genomförande och utvärdering av reformer. Det har varit tätt mellan olika reformer under 2000-talen och det finns en risk att beslutade reformer inte alltid har hunnit ge resultat innan en ny reform ska genomföras.

Lärarna i avhandlingen hade svårt att motivera sina bedömningsbeslut och peka på samband mellan syften, innehåll, förmågor och val av metoder. Forskning och bedömningslitteratur lyfter dock fram sådana samband, där olika metoder är mer lämpliga för att testa olika typer av kunnande. Samtidigt pekade lärarna på en del samband. Dessa hade de främst funnit utifrån egen erfarenhet, då de saknade utbildning i bedömning. Lärares bedömarkompetens skulle således kunna utvecklas i detta avseende och utbildning i bedömning har också visat sig bidra till att utveckla lärares bedömningspraktiker (Koloi-Keaikitse, 2016). Sammantaget pekar detta på vikten av utbildning om bedömning, både inom lärarutbildningen och vidareutbildning av verksamma lärare. De lärare som har påbörjat sin utbildning efter 2011 har fått sådan undervisning, vilket bör göra dem bättre rustade för detta arbete. Utbildningen i bedömning bör innehålla kunskaper om både viktiga bedömningsteoretiska begrepp och om hur ändamålsenliga bedömningar kan utformas.

På detta sätt kan lärare få ökad såväl konceptuell som praxeologisk kompetens (Pastore & Andrade, 2019). Det skulle kunna ge lärare en bredare bedömningsrepertoar, vilket också skulle möjliggöra en större agens. En sådan utbildning skulle också kunna bidra till att utveckla lärares professionella språk om bedömning, vilket kan hjälpa lärare att reflektera om sin bedömningspraktik.

Variationen mellan lärare, vilken avhandlingen har visat, kan vara problematisk utifrån likvärdighet och rättvisa. Elever riskerar att lära sig olika kunskapsinnehåll och förmågor och att bedömas på olika grunder. Ett professionellt språk skulle också kunna bidra till att utveckla professionella samtal mellan lärare, vilket skulle kunna minska problem med likvärdighet och rättvisa (Jönsson & Thornberg, 2014). Samtidigt lyfter lärarna i denna avhandling vikten av att enskilda lärare varierar sina bedömningar, för att kunna anpassa dessa efter olika elever och elevgruppers intressen och behov.

En viktig kvalitet i bedömningar är validitet (Dolin m.fl., 2018). Det finns dock ett behov av att anpassa begreppet validitet utifrån förutsättningarna i skolan (Bonner, 2013). Denna studie bidrar med insikter om denna problematik, bland annat då lärare ofta har många olika syften med sina bedömningar, vilka de behöver balansera. Lärare behöver en medvetenhet om behovet av att utforma ändamålsenliga uppgifter i relation till bedömningens syften, men de behöver också prioritera mellan sina syften.

Det är många olika beslut i en bedömningsprocess och dessa sker inom olika kontexter. En stark performativ diskurs tycks dock tendera att lärare främst utvecklar en summativ bedömningspraktik med betoning på att ha underlag för att sätta betyg. Detta riskerar att fokus flyttas från lärande till att bli en instrumentell praktik (Grettve m.fl., 2014; Löfgren m.fl., 2021; Olovsson, 2015). Samtidigt visar avhandlingen att lärare i samhällskunskap har stort frirum och agens att utforma sina bedömningar. Studien bidrar till en ökad förståelse för komplexiteten i lärares bedömningsarbete och om lärares agens, vilket kan bidra till att utveckla lärares reflektion om sin bedömningspraktik. Kanske kan de också upptäcka ett större frirum som möjliggör en utveckling av bedömningspraktiken med ett större fokus på lärande.

8 Summary

Assessments are crucial to teachers' work and the school system, influencing individuals' opportunities in both upcoming education and future life choices. However, there are relatively few studies on teachers' assessment practices regarding what is assessed, how it is assessed, the basis for teachers' assessment decisions, and what may influence these decisions. Existing research on teachers' assessment practices often relies on teachers' self-assessments through surveys (McMillan, 2013). This study is based on Civics teachers' own assessment tasks and interviews, providing a more comprehensive understanding of teachers' assessment practices and the rationale behind them. Consequently, it sheds light on a central aspect of teachers' work.

The aim is to investigate the assessment practices of Civics teachers in high school, focusing on purposes, content, skills, and methods, as well as factors influencing these practices.

1. What purposes of assessments do Civics teachers express?
2. What content and skills do Civics teachers assess, and with which methods?
3. What connections between the purposes, content, skills, and methods of assessments do Civics teachers express, and how do these relate to validity?
4. How do Civics teachers reason about factors influencing their decisions regarding assessments, and what agency do they express?

Background

In Sweden, there has been a long history of assessments, and grades being used for selection in education and employment (Lundahl, 2020). This makes teachers' assessments crucial for students as they impact their future opportunities. Furthermore, teachers' assessments have been predominantly based on written exams, primarily testing factual knowledge. Grading in Sweden has shifted from norm-referenced assessment to criterion-referenced, signifying a move from primarily assessing factual knowledge to also evaluating understanding and reasoning skills.

Regardless of the grading system, there have also been issues with equity. The latest grading reforms have introduced a performative discourse, which has had a significant impact on teachers' assessment practices within the Swedish testing culture, limiting them within this discourse (Wahlström & Sundberg, 2015). At the same time, a formative discourse evolved that underscores the use of assessment to enhance students' learning. Teachers need to consider both these different aspects

when shaping their assessments. Still, in the end it is the teachers who construct their assessments and assign grades, with some degree of professional autonomy.

Civics in Sweden

In Swedish upper secondary education, students choose between vocational and higher education preparatory programs. The syllabi of all courses consist of subject aim, core content, and grading criteria. Civics is mandatory for all programs. The courses Civics 1a1/1b, cover various topics such as democracy, human rights, and economics. While both courses share content, the course 1a1 in vocational programs is half as extensive, excluding political ideologies, national economics, and social science. Grading criteria are generally broad, with specific criteria on human rights and personal finance. Despite perceptions of an open and flexible subject (Öberg & Bäckström, 2021), studies suggest a core in political science and economics (Odenstad, 2011; Olsson, 2016).

The character of Civics is fostering citizenship, emphasizing civic education, reasoning skills, and analysis (Möllenborg, 2023; Sandahl et al., 2022; Wahlström et al., 2020). Wahlström et al. claim that generic competencies are emphasized more than subject-specific skills, while Möllenborg found that teachers used skills such as analysis in subject-specific tasks. Teachers therefore seem to perceive those skills differently in Civics than in other subjects.

Theoretical background

This thesis explores what teachers assess in their assessment tasks, encompassing various forms of knowledge. In the analyses of knowledge forms, Bloom's revised taxonomy was used in sub-studies I and III, along with interviews. Additionally, one research question pertains to validity.

The purpose involves examining teachers' reasoning about factors influencing their assessment practices. Throughout the sub-studies, various theories and models contribute to understanding teachers' assessment practices. In the first sub-study, frame factor theory is used to analyse factors framing teachers' work. In sub-study II, a model of teachers' assessment decision-making is applied. Sub-study III utilizes an ecological perspective on teacher agency. While these theories share similarities, they also complement each other, shedding light on factors influencing teachers' assessment practices.

Bloom's Revised Taxonomy

The thesis employs Bloom's revised taxonomy (Anderson et al., 2001) to categorize knowledge forms tested in teachers' assessment tasks. The taxonomy, with dimensions of cognitive processes and knowledge types, aids in detailed categorization of questions in sub-study I. In sub-study III, a simplified form of the taxonomy is used to categorize knowledge in teachers' submitted assessment tasks.

Validity

Two of the thesis questions address teachers' assessment purposes and the relationship between purposes, tested knowledge, and methods, making validity crucial. Validity in assessments is seen as essential for high-quality evaluations aligned with their intended purpose. Definitions of validity vary, but Messick's (1989) universal definition is used in the thesis. According to Messick, validity concerns the extent to which empirical evidence supports interpretations and uses of assessment results.

One aspect involves evidence for interpretation and use, meaning that the foundations must be relevant to the task being assessed, relating to the content and skills to be evaluated. Validation entails assessing the extent to which interpretations and uses are reasonable and relevant based on the assessment results. The second aspect concerns the consequences of assessments, both in terms of how the interpretation of the assessment can influence perceptions and values, such as determining important knowledge, and the social consequences the use of the assessment may have on motivation and self-esteem. However, validity can never be completely proven: an assessment can be more or less valid.

Threats to validity are construct underrepresentation and construct-irrelevant variance. Irrelevant matters are different forms of bias. It is important that an assignment is constructed so that it is representative of both the content and the abilities to be assessed, but also that it does not contain things that can distort the results or complicate the interpretation of the students' performance. Also, convergent evidence is needed, which means that the interpretation of an assessment needs to be consistent with other measures of the same skill being tested. This means variation is important.

Validity is crucial to both summative and formative assessments, as appropriate evidence is important for both grading, giving effective formative feedback, and adapting teaching.

Theories about factors influencing teachers' work

Frame factor theory (Imsen, 1999; Lundgren, 1989) explains classroom teaching processes in terms of various factors. Frames are factors that enable some actions in the classroom, but limit others. These can be pedagogical, administrative, resource-related, organizational-related, or student-related frames. The theory does not explicitly account for teachers as actors but emphasizes the interplay between frames, teaching processes, and outcomes. Frames may also be internal, how the teachers perceive the outer frames, which affects the teachers' perceived freedom within the frames.

The decision-making model (McMillan, 2003) illustrates that teachers' assessment practices result from internal factors (knowledge, perceptions, values) and external factors (policy, parents, external assessments). Teachers navigate tensions between these factors in deciding assessment forms, task design, and alignment with learning objectives.

Teachers may achieve agency, an ability to act, in different situations (Priestley et al., 2015). The ecological perspective on teacher agency considers the interaction between individual capacity and contextual conditions, encompassing historical, present, and future dimensions. Teachers' agency is influenced by their experiences, professional engagement, and the broader school context. A broad repertoire of methods and subject knowledge enriches teacher agency. The model introduces iterative (historical), projective, and practical-evaluative dimensions, emphasizing the importance of teachers' ambitions, historical influences, and context.

Teachers achieve agency when they have the opportunity to choose between different options in a given situation and are able to evaluate which option is most desirable in relation to the purposes of the practice in which they are acting (Priestley et al., 2015). Few alternatives will limit agency. Agency is absent if there are no courses of action or if teachers simply follow routines or habits without considering different options. This is generally true in teachers' work, as well as in the assessment process. The ecological perspective on teacher agency offers a comprehensive analytical model, emphasizing teacher agency, and will be primarily used in the subsequent discussion, with connections to other theoretical perspectives.

Earlier research

Research has indicated that teachers may have different purposes for their assessments (Harris & Brown, 2009; Newton, 2017), with both summative and formative purposes being central (Dolin et al., 2018).

Historically, teachers in various subjects have predominantly used written exams focusing on factual knowledge (Frey & Schmitt, 2010; Mattson, 1989; Moss, 2013). Written tests appear to be more dominant in higher age groups, while other forms are more common in younger ages, although there is some variation (Brookhart, 2013; Zhang & Burry-Stock, 2003). Practices may also differ between countries (Kippers et al., 2018; Osman, 2021).

The character of the subject seems to be relevant, at least in grading assessments (Biberman-Shalev et al., 2011; Prøitz, 2013; Zhang & Burry-Stock, 2003), but how the subject's character affects individual tasks is less explored. Information regarding assessments in Civics is scarce (Torrez and Claunch-Lebsack 2014). Actual practices of Civics teachers appear to deviate from the character of Civics, relying on objective tests that primarily measure factual knowledge, instead of performance assessments (Brookhart & Durkin, 2003; Curry & Smith, 2017; Mattson, 1989; Odenstad, 2011).

Various factors influence teachers' assessment practices. Several studies have highlighted that teachers often lack training in assessment, with prior experiences being more significant (Amani et al., 2021; Campbell, 2013; Forsberg & Wermke, 2012). Other aspects, such as teachers' preferences and emotions, can also have an impact (Edwards, 2020; Vanlommel et al., 2017). The performative discourse has proven to influence teachers' agency, while teachers may also misinterpret the requirements (Olovsson, 2015; Priestley et al., 2015).

Teachers' assessment practices have consequences, affecting students' learning, motivation, emotions, and self-confidence, as indicated by research (Brookhart & Durkin, 2003; Hirsh, 2020; Löfgren et al., 2019; Pekrun et al., 2023). Furthermore, these practices have implications for equity, fairness, and validity.

Methodology

Two different data collections were conducted, one for the first sub-study and another for the other two. In both data collections, semi-structured interviews were conducted, but with different focuses and questions. The first data collection involved a preliminary study where 135 teachers, constituting all Civics teachers in a medium-sized county in Sweden, were asked to send two tests (one about how Sweden is governed and one about social economics). Of these, 21 teachers submitted tests. The tests were analysed using Bloom's revised taxonomy, categorizing teachers into three groups based on the proportion of remembering questions (<50%, 50-80%, and >80%). Two teachers from each category, a total of six, were selected for follow-up interviews. Interviews were conducted in

autumn 2009 at the teachers' workplaces. The second data collection for sub-studies II and III involved sending inquiries to 108 civics teachers in 17 different municipalities of varying sizes, and in three different regions. Sixteen teachers submitted assignments, displaying a wide range of tasks. Thirteen of these teachers were interviewed in spring 2019 at their workplaces.

In terms of analysis, sub-study I used a form of qualitative text analysis (Boréus & Bergström, 2018), combining submitted tests with teachers' statements to identify knowledge forms tested in the questions. Teachers' statements about their tests were also analysed using Bloom's revised taxonomy. Interviews were also analysed for expressions of frame factors influencing their assessment work. Sub-studies II and III utilized thematic analysis to identify various purposes of teachers' assessments, and to analyse the content, skills, and methods used in assessments. The thematic analysis process involved multiple rounds of listening and reading through interviews and teachers' tasks, followed by inductive coding and the generation of themes, with several sub-themes.

Following ethical guidelines (Vetenskapsrådet, 2017), participants were informed about the study's objectives, and what voluntariness of participation and anonymity were guaranteed.

Semi-structured interviews were chosen to explore factors influencing teachers' assessment decisions, providing a flexible yet structured approach. The researcher's background as a former secondary school teacher in Civics was considered both an advantage and a potential limitation. The study's qualitative nature enables a richness of the description of empirical data, balancing the need for transparency and reducing complexity.

Summary of the sub-studies

The thesis consists of three sub-studies, a licentiate thesis and two articles, which are summarized below.

Study I – What is tested? (Licentiate Thesis)

The objective of this study was to explore secondary school Civics teachers' assessment practices. The primary focus was to evaluate, utilizing Bloom's revised taxonomy, the types of knowledge examined in written tests. The analyses indicated a predominant emphasis on testing factual and conceptual knowledge, while procedural knowledge is evaluated through alternative means, primarily essays. However, meta-cognitive knowledge is absent from the study. Unlike prior research, this study revealed that teachers assessed a diverse range of cognitive processes,

with 50 to 90% of questions measuring the category to *remember*, and fewer questions measuring to *apply* and *create*. Although teachers demonstrate a good understanding of grading criteria and the curriculum, discrepancies exist between this understanding and their actual testing practices, raising concerns about test validity.

Given that the tests primarily serve summative purposes, variations in teachers' practices contribute to issues of grade consistency. Discrepancies in practices may be explained by frame factors, especially administrative factors such as working hours and schedules. Adequate time is crucial for teachers to prepare and mark tests. Also, students prefer written tests, prompting teachers to accommodate these preferences. Nevertheless, the impact of these factors is contingent on teachers' freedom. While most teachers possess the knowledge and intent to create valid tests, they require sufficient time for preparation and grading, along with the flexibility to extend lessons for additional writing time.

Study II – Balancing diverse purposes in Civics teachers' assessment decision making (Article)

The purpose of this study is to describe the various purposes of assessments as expressed by Swedish Civics teachers in high school. The study revealed that high school civics teachers in Sweden have diverse and numerous purposes for their assessments, aligning with prior research. These purposes can be categorized into four main themes.

The first theme, 'implementing policy', pertains to external factors. Teachers expressed intentions to meet curriculum and syllabus requirements, particularly the knowledge criteria. Providing a basis for grading was a significant assessment purpose, though challenges arose with some content-specific criteria. The second theme involves purposes related to 'handling administration', also an external factor. Teachers aimed to fit assessments within working hours, class time, and schedules. The third theme, 'improving learning', primarily relates to internal factors. Teachers expressed a personal desire to foster learning, emphasizing formative assessment through feedback, peer assessment, and self-assessment. Their objectives included deepening understanding, preparing students for future studies, and imparting knowledge of current events. The final theme is 'pulling for students', also related mainly to internal factors. Teachers sought to provide students with influence over assessments, allowing them to choose timing and formats to increase success chances. Accommodations were made based on individual needs, group, class, and program requirements.

Balancing these diverse purposes, along with tensions between formative and summative assessments, are challenging for teachers. Exter-

nal pressures, such as the emphasis on grading in Swedish policy documents, contribute to these tensions. The study suggests that a simplified validation process focusing on primary assessment purposes could aid teachers in managing the complexities of their assessment practices effectively. Understanding the multifaceted nature of teachers' assessment processes equips teachers to make informed decisions in their daily practices.

Study III – Civics teachers' assessment practices in Swedish upper secondary schools. A qualitative study (Article)

The aim was to qualitatively explore Civics teachers' assessment decisions regarding content, skills, and methods in Swedish upper secondary schools. The Civics teachers expressed that they could achieve agency in their assessment work. They have a freedom to choose what content to assess and with which methods, a freedom that the teachers in the study applied using a variety of methods and different focuses on content. However, if the variation between teachers becomes too great, there is a risk that the pupils will be assessed on different grounds. Since assessments affect students' learning strategies, they are thus at risk of developing different learning outcomes, which can lead to problems with equitable assessment. At the same time, this freedom can enable teachers to adapt assessments based on the students' conditions and interests, but also based on the teachers' own interests, which can make the work more meaningful.

The Civics teachers in the study had a more varied assessment practice that, to a greater extent, focuses on reasoning abilities, compared to older studies, in which written tests measuring to remember factual knowledge dominate. This indicates that Civics teachers' assessments today are more valid in relation to the curriculum.

Teachers need to balance between a summative and a formative discourse. They also need to balance this and their own preferences with administrative requirements in the form of limited working hours and schedules. Teachers' agency could be greater if they were less bound by subject traditions, but also through education in assessment, which could provide a more extensive assessment repertoire. This, in turn, would provide increased opportunities for higher validity as teachers would be better equipped to choose relevant methods depending on the content and abilities to be assessed.

Discussion

The conclusion is that even though the Civics teachers had varied purposes, their assessment practice was strongly summative oriented. Formative purposes influenced their design of assessment tasks to a lesser extent. A variety of methods enables assessments to be adapted to students' different needs, helping them to succeed. At the same time, there is a risk that pupils develop different knowledge and skills and that they are assessed on different grounds, which could affect equality and fairness. There are tensions between the teachers' different purposes, which they need to balance in their professional practice. Different purposes also complicate the question of validity because teachers' tasks need to be relevant in relation to their purposes (Kane & Wools, 2020; Newton & Shaw, 2016). However, different contextual conditions give teachers different agency to deal with these tensions.

The steering documents both limit and enable teacher agency in the work with assessments. The teachers expressed greater agency in relation to the choice of content and methods, but a narrower agency when it comes to decisions about skills because of the requirement to have evidence for all grading criteria. Teachers testing different content can lead to students learning different content.

The openness and flexibility of the subject enables Civics teachers a high degree of agency regarding their choice of content and methods. The character of the subject affects the teachers' assessment practices regarding content, in the form of actualization, what skills are assessed due to the emphasis on reasoning tasks and source criticism, and some decisions about methods. At the same time, a subject tradition also affects the content that teachers assess.

Validity does not seem to be something the teachers have reflected on. This points to a risk that teachers do not always design their tasks in an appropriate way, not measuring what they intend to. Training for both pre-service and in-service teachers could increase their understanding of validity, as well as their competence in constructing valid assessments based on different purposes, i.e. their conceptual and praxeological assessment literacy (Pastore & Andrade, 2019). At the same time, Civics teachers' assessments tended to be more valid for summative purposes, but not as clearly in relation to formative purposes.

Additionally, it is important to be aware of the consequences of assessment practices for students. There are some challenges for teachers in dealing with negative consequences of assessments and instead strengthening positive effects to increase students' motivation and self-confidence, and to promote learning of different skills.

The overall picture is that the Civics teachers make an effort to construct tasks that will test the abilities that are expressed in the grading criteria, while the central content is not emphasized in the same way. The conclusion is that teachers are willing and able to make valid assessments under favourable contextual conditions and with increased knowledge about the construction of appropriate assessments. However, there is also a need for an awareness that assessments cannot be valid for all different purposes at the same time, but that teachers need to prioritize which purposes are most important (Kane & Wools, 2020). They also need to consider the impact of assessments by minimizing the negative consequences and reinforcing the positive ones. However, this may conflict with validity in relation to some purposes, such as grading, while other purposes such as promoting learning and helping students succeed, leading to more positive consequences. This can also be problematized as a too strong emphasis on these latter purposes could lead to problems regarding equal assessment. Teachers have to balance all of these different requirements in their assessment practice, thus confirming that assessment is a difficult and complex task, but also an important one, where teachers have the opportunity for great agency and professionalism.

Limitations and implications

This study contributes to understanding the assessment practices of Swedish Civics teachers in upper secondary schools. However, knowledge about how teachers conduct assessments is still limited, both in Sweden and internationally, and thus more research is needed. The dissertation employs qualitative methods, including interviews and analyses of assessment tasks from a relatively small number of teachers. While the results cannot be quantitatively generalized, they can be qualitatively generalized (Larsson, 2010). The strategic selection of informants aims for a broad representation of teachers with diverse backgrounds and contexts, enhancing the study's qualitative generalization. The study's mixed method, incorporating both collected artifacts and interviews, strengthens its findings. The use of only written task instructions is a limitation, and exploring oral tasks could provide additional insights. The study suggests the need for similar research in other subjects and at different educational levels.

The research identifies a weak framing of the subject of social studies, echoing previous literature, and encourages further analysis of teachers' assignments to better understand the emphasized knowledge in the subject. Studies comparing Civics teaching and assessment in vocational versus higher education preparatory programs, as well as in other subjects, are recommended. The study observes limited use of digital assess-

ments, prompting interest in exploring the impact of digital tools on teachers' assessment practices, considering potential changes in task formats, or increased use of self-scoring questions. The ongoing debate on AI's influence on assessments, including the risk of cheating, underscores the need for investigation into the evolving landscape of digital assessment tools.

Assessment is a crucial task for teachers, impacting individual students and society. This thesis contributes insights into this area. The study indicates a slow shift in assessment practices in Civics, meaning that it takes a long time for reforms to take effect; this needs to be considered by policy makers.

Teachers' decisions about assessments appear at times unreflective, indicating a need for improved assessment competence and training for teachers. The study advocates ongoing teacher education to enhance both conceptual and practical competence, fostering a professional language around assessment. Validity in assessments is crucial, and teachers must balance multiple purposes, emphasizing the need for thoughtful task design aligned with assessment purposes.

This dissertation underscores the need for adapting the concept of validity to school contexts and highlights the complexity of decision-making in the assessment process. Despite teachers having significant agency in designing assessments, a strong performative discourse tends to drive a summative assessment focus, instead of promoting learning. The study contributes to an increased understanding of the complexity of teachers' assessment work and teacher agency, contributing to teachers' reflection on their assessment practice and perhaps helping them to discover a larger freedom that enables a development of assessment practices with a greater focus on learning.

Referenser

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20–34. https://doi.org/10.1080/0969594X.2012.736364

Amani, J., Kitta, S., Kapinga, O. S. & Mbilinyi, C. (2021). Secondary school teachers' knowledge on procedures for constructing quality classroom tests in Tanzania. *Üniversitepark Bülten*, *10*(1), 40–54. https://doi.org/10.22521/unibulletin.2021.101.3

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. P., Raths, J. & Wittrock, M. C. (Red.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Arevik, S. & Hartzell, O. (2015). *Att göra tänkande synligt: En bok om begreppsbaserad undervisning* (2 uppl.). Liber.

Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal*, *25*(2), 238–259. https://doi.org/10.1080/09585176.2013.874952

Barton, K. C. & Avery, P. G. (2016). Research on social studies education: Diverse students, settings, and methods. I D. H. Gitomer & C. A. Bell (Red.), *Handbook of research on teaching* (5 uppl., s. 985–1038). American Educational Research Association. http://ebooks.aera.net/HRTCH16

Becevic, S. (2011). *Klassrumsbedömning i matematik på gymnasieskolans nivå* [Licentiatuppsats, Linköpings Universitet]. Linköping studies in science and technology, 1502. https://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A439814&dswid=-6887

Berg, M. & Persson, A. (2020). Graderande granskning och förklarade glapp: Svensk historie- och samhällskunskapsdidaktisk forskning om lärares bedömningspraktik 2009– 2019. *Nordidactica. Journal of Humanities and Social Science Education*, *10*(1), 18–44.

Bergh, A. & Wahlström, N. (2018). Conflicting goals of educational action: A study of teacher agency from a transactional realism perspective. *The Curriculum Journal*, *29*(1), 134–149. https://doi.org/10.1080/09585176.2017.1400449

Bernmark-Ottosson, A. (2009). Samhällskunskapslärare. I B. Schüllerqvist & C. Osbeck (Red.), *Ämnesdidaktiska insikter och strategier: Berättelser från gymnasielärare i samhällskunskap, geografi, historia och religionskunskap* (s. 33–82). Karlstad University Press.

Bernstein, B. (2003). *Class, codes and control Vol. 1 Theoretical studies*

towards a sociology of language. Routledge.
Betygsberedningen. (2008). *En ny betygsskala* (Utredning Ds 2008:13; Departementsserien). Utbildningsdepartementet. https://www.regeringen.se/rattsliga-dokument/departementsserien-och-promemorior/2008/02/ds-200813/
Betygsutredningen. (2020). *Bygga, bedöma, betygssätta – betyg som bättre motsvarar elevernas kunskaper* (Utredning SOU 2020:43; Statens offentliga utredningar). Regeringskansliet. https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2020/08/sou-202043/
Biberman-Shalev, L., Sabbagh, C., Resh, N. & Kramarski, B. (2011). Grading styles and disciplinary expertise: The mediating role of the teacher's perception of the subject matter. *Teaching and Teacher Education*, *27*(5), 831–840. https://doi.org/10.1016/j.tate.2011.01.007
Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 161–177. https://doi.org/10.1080/0969594X.2014.999643
Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 215–232. https://doi.org/10.1080/09695941003696016
Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 87–106). SAGE Publications. https://doi.org/10.4135/9781452218649
Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. I G. Brown & L. Harris (Red.), *Handbook of human and social conditions in assessment* (s. 21–39). Routledge.
Bonner, S. M., Torres Rivera, C. & Chen, P. P. (2018). Standards and assessment: Coherence from the teacher's perspective. *Educational Assessment, Evaluation and Accountability*, *30*(1), 71–92. https://doi.org/10.1007/s11092-017-9272-2
Boréus, K. & Bergström, G. (2018). *Textens mening och makt: Metodbok i samhällsvetenskaplig text- och diskursanalys* (4 uppl.). Studentlitteratur.
Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
Braun, V. & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE Publications.
Braun, V. & Clarke, V. (2022). *Thematic analysis: A practical guide*.

SAGE Publications.
Brookhart, S. M. (2009). Mixing it up: Combining sources of classroom achievement information for formative and summative purposes. I H. Andrade & G. J. Cizek (Red.), *Handbook of formative assessment [eBook]* (s. 279–296). Taylor & Francis Group. https://doi.org/10.4324/9780203874851
Brookhart, S. M. (2013). Grading. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 256–271). SAGE Publications. https://doi.org/10.4135/9781452218649.n15
Brookhart, S. M. & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, *16*(1), 27–54. https://doi.org/10.1207/S15324818AME1601_2
Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301–318. https://doi.org/10.1080/0969594042000304609
Bryman, A. (2016). *Social research methods* (5 uppl.). Oxford University Press.
Campbell, C. (2013). Research on teacher competency in classroom assessment. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 71–84). SAGE Publications. https://doi.org/10.4135/9781452218649.n5
Carlgren, I., Forsberg, E. & Lindberg, V. (2009). *Perspektiv på den svenska skolans kunskapsdiskussion*. Stockholms universitets förlag.
Chappuis, J. & Stiggins, R. (2019). *Classroom assessment for student learning: Doing it right—Using it well* (3 uppl.). Pearson.
Cimer, S. O. & Cimer, A. (2010). What teachers assess and its consequences. *Asia-Pacific Forum on Science Learning and Teaching*, *11*(2), 1–24.
Cizek, G. J. (2009a). An introduction to formative assessment: History, characteristics, and challenges. I H. L. Andrade & G. J. Cizek (Red.), *Handbook of formative assessment* (s. 3–17). Routledge. https://doi-org.e.bibl.liu.se/10.4324/9780203874851
Cizek, G. J. (2009b). Reliability and validity of information about student achievement: Comparing large-scale and classroom testing contexts. *Theory Into Practice*, *48*(1), 63–71. https://doi.org/10.1080/00405840802577627
Colnerud, G. & Granström, K. (2015). *Respekt för lärarprofessionen: Om lärares yrkesspråk och yrkesetik* (4 uppl.). Liber.
Cooper, A., Klinger, D. A. & McAdie, P. (2017). What do teachers need? An exploration of evidence-informed practice for classroom

assessment in Ontario. *Educational Research*, *59*(2), 190–208. https://doi.org/10.1080/00131881.2017.1310392
Curry, K. & Smith, D. (2017). Assessment practices in social studies classrooms: Results from a longitudinal survey. *Social Studies Research and Practice*, *12*(2), 168–181. https://doi.org/10.1108/SSRP-04-2017-0015
David, M. & Sutton, C. D. (2016). *Samhällsvetenskaplig metod*. Studentlitteratur.
Dolin, J., Black, P., Harlen, W. & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. I J. Dolin & R. Evans (Red.), *Transforming assessment* (s. 53–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-63248-3_3
Duncan, C. R. & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, *53*(1), 1–21. https://doi.org/10.11575/ajer.v53i1.5519
Edwards, F. (2020). The effect of the lens of the teacher on summative assessment decision making: The role of amplifiers and filters. *The Curriculum Journal*, *31*(3), 379–397. https://doi.org/10.1002/curj.4
Elwood, J. & Murphy, P. (2015). Assessment systems as cultural scripts: A sociocultural theoretical lens on assessment practice and products. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 182–192. https://doi.org/10.1080/0969594X.2015.1021568
Englund, T. (1986). *Samhällsorientering och medborgarfostran i svensk skola under 1900-talet*. Pedagogiska institutionen, Uppsala Universitet.
Forsberg, E. & Lindberg, V. (2010). *Svensk forskning om bedömning: En kartläggning*. Vetenskapsrådet.
Forsberg, E. & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education*, *38*(5), 741–758. https://doi.org/10.1080/19415257.2012.694369
Frey, B. B. & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal*, *5*(3), 107–117.
Gioka, O. (2009). Teacher or examiner? The tensions between formative and summative assessment in the case of science coursework. *Research in Science Education*, *39*(4), 411–428. https://doi.org/10.1007/s11165-008-9086-9
Gomez, M. D. C. (2017). *Teachers' Assessment and grading Practices in upper secondary Science Classrooms in Sweden: The Teachers' and Students' Perspectives*. [Doktorsavhandling, Lund University]. Education. https://www.lu.se/lup/publication/474c2b32-b070-489a-9772-a80bb573372f

Green, S. K., Johnson, R. L., Kim, D.-H. & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, *23*(7), 999–1011. https://doi.org/10.1016/j.tate.2006.04.042

Grettve, A., Israelsson, M. & Jönsson, A. (2014). *Att bedöma och sätta betyg: Tio utmaningar i lärarens vardag*. Natur & Kultur.

Gu, P. Y. (2021). An argument-based framework for validating formative assessment in the classroom. *Frontiers in Education*, *6*, 1–10. https://doi.org/10.3389/feduc.2021.605999

Guskey, T. R. & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, *26*(3), 303–320. https://doi.org/10.1080/0969594X.2018.1555515

Gustafsson, C. (1999). Ramfaktorer och pedagogiskt utvecklingsarbete. *Pedagogisk Forskning i Sverige*, *4*(1), 43–57.

Hadley, G. R. L. & Young, D. C. (2018). A glimpse inside: Considering the impact of curriculum outcomes and personal ideology on social studies pedagogy: A study summary. *Journal of International Social Studies*, *8*(1), 25.

Hargreaves, A. (1989). *Curriculum and assessment reform*. Open University Press.

Harris, L. R. & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 365–381. https://doi.org/10.1080/09695940903319745

Hartell, E. & Strimel, G. J. (2019). What is it called and how does it work: Examining content validity and item design of teacher-made tests. *International Journal of Technology and Design Education*, *29*(4), 781–802. https://doi.org/10.1007/s10798-018-9463-2

Heritage, M. (2013). Gathering evidence of student understanding. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 179–195). SAGE Publications. https://doi.org/10.4135/9781452218649.n11

Hirsh, Å. (2020). When assessment is a constant companion: Students' experiences of instruction in an era of intensified assessment focus. *Nordic Journal of Studies in Educational Policy*, *6*(2), 89–102. https://doi.org/10.1080/20020317.2020.1756192

Imsen, G. (1999). *Lärarens värld: Introduktion till allmän didaktik*. Studentlitteratur.

James, M. & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, *17*(2), 109–138. https://doi.org/10.1080/09585170600792712

Jansson, T. (2011). *Vad kommer på provet? Gymnasielärares provpraxis i samhällskunskap* [Licentiatuppsats, Karlstad Universitet]. Karlstad University Studies: Studier i de samhällsvetenskapliga ämnenas didaktik, 5. https://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A380888

Jansson, T. (2023). Civics teachers' assessment practices in Swedish upper secondary schools. A qualitative study. *Journal of Social Science Education, 22*(2). https://doi.org/10.11576/jsse-5938

Jansson, T. & Löfgren, H. (2022). Balancing diverse purposes in Civics teachers' assessment decision making. *Scandinavian Journal of Educational Research*, 1–15. https://doi.org/10.1080/00313831.2022.2148280

Jølle, L. & Skar, G. B. (2020). "Digging for gold" or "Sticking to the criteria": Teachers' rationales when serving as professional raters. *Scandinavian Journal of Educational Research*, *64*(2), 283–299. https://doi.org/10.1080/00313831.2018.1541821

Jönsson, A. (2023a). *Lärande bedömning* (6 uppl.). Gleerups.

Jönsson, A. (2023b). *Prov eller bedömning? Att tolka och använda provresultat och omdömen* (2 uppl.). Gleerups.

Jönsson, A. & Thornberg, P. (2014). Samsyn eller samstämmighet? En diskussion om sambedömning som redskap för likvärdig bedömning i skolan. *Pedagogisk Forskning i Sverige*, *4/5*, 386–402.

Kakoulidou, K. (2021). *Från A till F: Bedömning och betygssättning i historia på gymnasiet* [Licentiatuppsats, Linnéuniversitetet]. Linnaeus University Press: Lnu Licentiate, 36. https://lnu.diva-portal.org/smash/record.jsf?pid=diva2%3A1609072&dswid=-6968

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. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. https://doi.org/10.1080/0969594X.2015.1060192

Kane, M. T. & Wools, S. (2020). Perspectives on the validity of classroom assessments. I S. M. Brookhart & J. H. McMillan (Red.), *Classroom assessment and educational measurement* (s. 11–26). Routledge, Taylor & Francis Group.

Karlefjärd, A. (2011). *Att rymmas inom sitt friutrymme: Om samhällskunskapslärares tolkning, anpassning och undervisning* [Licentiatuppsats, Karlstad Universitet]. Karlstad University Studies: Studier i de samhällsvetenskapliga ämnenas didaktik, 12. https://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A395564&dswid=3785

Karlsson, A. (2011). *Samhällsguide, individualist och moderator:*

Samhällskunskapslärares professionella förhållningssätt i betygsättningsrelaterat arbete [Licentiatuppsats, Karlstad Universitet]. Karlstad University Studies: Studier i de samhällsvetenskapliga ämnenas didaktik, 13. https://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A395830&dswid=-4433

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L. & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, *75*, 199–213. https://doi.org/10.1016/j.tate.2018.06.015

Klapp, A. (2016). Vad mäter betygen. I C. Lundahl & M. Folke-Fichtelius (Red.), *Bedömning i och av skolan: Praktik, principer, politik* (2 uppl., s. 135–148). Studentlitteratur.

Klapp Ekholm, A. (2010). *Bedömning för lärande: En grund för ökat kunnande*. Stiftelsen SAF i samarbete med Lärarförbundet.

Klenowski, V. (2013). Towards improving public understanding of judgement practice in standards-referenced assessment: An Australian perspective. *Oxford Review of Education*, *39*(1), 36–51. https://doi.org/10.1080/03054985.2013.764759

Koloi-Keaikitse, S. (2016). Assessment training: A precondition for teachers' competencies and use of classroom assessment practices: Assessment training. *International Journal of Training and Development*, *20*(2), 107–123. https://doi.org/10.1111/ijtd.12072

Korp, H. (2006). *Lika chanser i gymnasiet? En studie om betyg, nationella prov och social reproduktion* [Doktorsavhandling, Malmö högskola]. Malmö Studies in Educational Sciences: Doctoral Dissertation Series, 24. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1404377&dswid=-3149

Korp, H. (2011). *Kunskapsbedömning: Vad, hur och varför?* Skolverket.

Kvale, S. & Brinkmann, S. (2014). *Den kvalitativa forskningsintervjun* (3 uppl.). Studentlitteratur.

Larsson, S. (2010). Om generalisering från kvalitativa studier. I I. Eriksson, V. Lindberg & E. Österlind (Red.), *Uppdrag undervisning: Kunskap och lärande* (s. 51–68). Studentlitteratur.

Larsson, S. (2005). Om kvalitet i kvalitativa studier. *Nordisk Pedagogik*, *25*(1), 16–35.

Lauvås, P., Jönsson, A. & Trägårdh, E. (2019). *Ren formativ bedömning: En ny bedömningspraktik*. Studentlitteratur.

Leong, W. S. (2014). Knowing the intentions, meaning and context of classroom assessment: A case study of Singaporean teacher's conception and practice. *Studies in Educational Evaluation*, *43*, 70–78. https://doi.org/10.1016/j.stueduc.2013.12.005

Lidar, M., Lundqvist, E., Ryder, J. & Östman, L. (2020). The transformation of teaching habits in relation to the introduction of grading and national testing in science education in Sweden. *Research in Science Education*, *50*(1), 151–173. https://doi.org/10.1007/s11165-017-9684-5

Lindberg, V. & Hirsh, Å. (2015). *Formativ bedömning på 2000-talet: En översikt av svensk och internationell forskning*. Vetenskapsrådet.

Lindblad, S., Linde, G. & Naeslund, L. (1999). Ramfaktorteori och praktiskt förnuft. *Pedagogisk Forskning i Sverige*, *4*(1), 93–109.

Lindblad, S. & Sahlström, F. (1999). Gamla mönster och nya gränser. *Pedagogisk Forskning i Sverige*, *4*(1), 73–92.

Linde, G. (2021). *Det ska ni veta! En introduktion till läroplansteori* (4 uppl.). Studentlitteratur.

Lindmark, T. (2013). *Samhällskunskapslärares ämneskonceptioner* [Doktorsavhandling, Umeå universitet]. Akademiska avhandlingar vid Pedagogiska institutionen, 108. https://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A663989&dswid=-2295

Lindqvist, S. & Hyltegren, G. (1995). *Att utveckla elevers tänkande: En teoretisk praktika för samhällsorienterande ämnen*. Almqvist & Wiksell.

Lindström, L. (2013). Pedagogisk bedömning. I L. Lindström, V. Lindberg & A. Pettersson (Red.), *Pedagogisk bedömning: Att dokumentera, bedöma och utveckla kunskap* (3 uppl., s. 9–27). Liber.

Looney, A., Cumming, J., van Der Kleij, F. & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090

Lumadi, M. W. (2013). Challenges besetting teachers in classroom assessment: An exploratory perspective. *Journal of Social Sciences*, *34*(3), 211–221. https://doi.org/10.1080/09718923.2013.11893132

Lundahl, C. (2020). Om bedömning. Att veta vad andra vet. I U. P. Lundgren, R. Säljö & C. Liberg (Red.), *Lärande, skola, bildning. Grundbok för lärare* (5 uppl., s. 547–586). Natur & Kultur.

Lundahl, C., Hultén, M., Klapp, A. & Mickwitz, L. (2015). *Betygens geografi. Forskning om betyg och summativa bedömningar i Sverige och internationellt: Delrapport från skolforsk-projektet*. Vetenskapsrådet.

Lundahl, C., Hultén, M. & Tveit, S. (2016). *Betygssystem i internationell belysning*. Skolverket.

Lundgren, U. P. (1989). *Att organisera omvärlden: En introduktion till*

läroplansteori (4 uppl.). Utbildningsförlaget på uppdrag av Gymnasieutredningen.
Löfgren, H., Alm, F., Markström, A.-M., Hultén, M., Jönsson, A. & Lundahl, C. (2021). *Betyg i årskurs 4: Ett forskningsunderlag om bedömningspraktikerna på skolor som deltagit i försöksverksamhet med tidig betygssättning*. Skolverket.
Löfgren, R., Löfgren, H. & Lindberg, V. (2019). Pupils' perceptions of grades: A narrative analysis of stories about getting graded for the first time. *Assessment in Education: Principles, Policy & Practice*, *26*(3), 259–277. https://doi.org/10.1080/0969594X.2019.1593104
Markus, K. A. (2016). Alternative vocabularies in the test validity literature. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 252–267. https://doi.org/10.1080/0969594X.2015.1060191
Mattson, H. (1989). *Proven i skolan sedda genom lärarnas ögon* [Licentiatuppsats, Umeå universitet]. Licentiatavhandlingar vid Pedagogiska institutionen, Umeå universitet, 4.
Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process — the assessment episode. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 381–400. https://doi.org/10.1080/0969594970040305
McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x
McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34–43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x
McMillan, J. H. (2013). Why we need research on classroom assessment. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 2–16). SAGE Publications. https://doi.org/10.4135/9781452218649.n1
McMillan, J. H. (2018). *Classroom assessment: Principles and practice that enhance student learning and motivation* (7 uppl.). Pearson.
McMillan, J. H. & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
McMillan, J. H. & Workman, D. (1999). *Teachers' classroom assessment and grading practices: Phase I and II*. Metropolitan Educational Research Consortium.
Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, *120*(2), 285–296.

Messick, S. (1989). Validity. I R. L. Linn (Red.), *Educational measurement* (3 uppl., s. 13–103). American Council on Education & Macmillan.

Moss, C. M. (2013). Research on classroom summative assessment. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 235–255). SAGE Publications. https://doi.org/10.4135/9781452218649.n14

Måhl, P. (1991). *Betyg—Men på vad? En bok om kunskapssyn och prov*. HLS (Högskolan för lärarutbildning).

Möllenborg, E. (2023). *Diskurser och dilemman i gymnasieskolans samhällskunskapsundervisning* [Doktorsavhandling, Linnaeus University]. Linnaeus University Dissertations, 499. https://lnu.diva-portal.org/smash/record.jsf?pid=diva2%3A1790966&dswid=1731

Newton, P. E. (2017). There is more to educational measurement than measuring: The importance of embracing purpose pluralism. *Educational Measurement: Issues and Practice*, *36*(2), 5–15. https://doi.org/10.1111/emip.12146

Newton, P. E. & Shaw, S. D. (2014). *Validity in educational & psychological assessment*. SAGE Publications.

Newton, P. E. & Shaw, S. D. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 178–197. https://doi.org/10.1080/0969594X.2015.1037241

Näsström, G. (2008). *Measurement of alignment between standards and assessment* [Doktorsavhandling, Umeå universitet]. Academic dissertations at the department of Educational Measurement, 3. https://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A142244&dswid=3357

Odenstad, C. (2011). *Prov och bedömning i samhällskunskap: En analys av gymnasielärares skriftliga prov* [Licentiatuppsats, Karlstad Universitet]. Karlstad University Studies: Studier i de samhällsvetenskapliga ämnenas didaktik, 4. https://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A379242&dswid=8312

Olovsson, T. G. (2015). *Det kontrollera(n)de klassrummet: Bedömningsprocessen i svensk grundskolepraktik i relation till införandet av nationella skolreformer* [Doktorsavhandling, Umeå universitet]. Umeå Studies in the Educational Sciences, 6. https://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A811225&dswid=9842

Olsson, R. (2016). *Samhällskunskap som ämnesförståelse och undervisningsämne: Prioriteringar och nyhetsanvändning hos fyra gymnasielärare* [Doktorsavhandling, Karlstads universitet].

Karlstad University studies, 9. https://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A902572&dswid=6888
Oo, C. Z., Alonzo, D. & Davison, C. (2021). Pre-service teachers' decision-making and classroom assessment practices. *Frontiers in Education*, *6*, 1-12. https://doi.org/10.3389/feduc.2021.628100
Oscarson, M. & Apelgren, B. M. (2011). Mapping language teachers' conceptions of student assessment procedures in relation to grading: A two-stage empirical inquiry. *System*, *39*(1), 2–16. https://doi.org/10.1016/j.system.2011.01.014
Osman, S. (2021). Basic school teachers' assessment practices in the Sissala east municipality, Ghana. *European Journal of Education Studies*, *8*(7). https://doi.org/10.46827/ejes.v8i7.3801
Pastore, S. & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003
Pekrun, R., Marsh, H. W., Suessenbach, F., Frenzel, A. C. & Goetz, T. (2023). School grades and students' emotions: Longitudinal models of within-person reciprocal effects. *Learning and Instruction*, *83*, 1–13. https://doi.org/10.1016/j.learninstruc.2022.101626
Penuel, W. R. & Shepard, L. A. (2016). Assessment and teaching. I D. H. Gitomer & C. A. Bell (Red.), *Handbook of research on teaching* (5 uppl., s. 787–850). American Educational Research Association. https://doi.org/10.3102/978-0-935302-48-6_12
Pettersson, D. & Wester, A. (2020). Skolan och världen— Internationella kunskapsmätningar. I U. P. Lundgren, R. Säljö & C. Liberg (Red.), *Lärande, skola, bildning*. *Grundbok för lärare* (5 uppl., s. 491–518). Natur & Kultur.
Priestley, M., Biesta, G. & Robinson, S. (2015). *Teacher Agency: An Ecological Approach*. Bloomsbury Academic.
Prop. 2009/10:89. *Bäst i klassen – en ny lärarutbildning*. https://www.regeringen.se/rattsliga-dokument/proposition/2010/02/prop.-20091089
Prøitz, T. S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*, *4*(3), 555–575. https://doi.org/10.3402/edui.v4i3.22629
Rahman, Md. M. (2018). Exploring Teachers Practices of Classroom Assessment in Secondary Science Classes in Bangladesh. *Journal of Education and Learning*, *7*(4), 274–283. https://doi.org/10.5539/jel.v7n4p274
Roos, B. & Hamilton, D. (2005). Formative assessment: A cybernetic viewpoint. *Assessment in Education: Principles, Policy & Practice*, *12*(1), 7–20. https://doi.org/10.1080/0969594042000333887
Rosenlund, D. (2016). *History education as content, methods or*

orientation? A study of curriculum prescriptions, teacher-made tasks and student strategies [Doktorsavhandling]. Malmö högskola.
Ryder, J., Lidar, M., Lundqvist, E. & Östman, L. (2018). Expressions of agency within complex policy structures: Science teachers' experiences of education policy reforms in Sweden. *International Journal of Science Education*, *40*(5), 538–563. https://doi.org/10.1080/09500693.2018.1435921
Samuelsson, J. (2017). Betygshistorik. I K. Nordgren, C. Odenstad & J. Samuelsson (Red.), *Betyg i teori och praktik: Ämnesdidaktiska perspektiv på bedömning i grundskola och gymnasium* (3 uppl., s. 21–37). Gleerups.
Samuelsson, J., Brismark Östlin, A. & Löfgren, H. (2018). *Papperspedagoger – lärares arbete med administration i digitaliseringens tidevarv* (Rapport 2018:5). IFAU - Institutet för arbetsmarknads- och utbildningspolitisk utvärdering. https://www.ifau.se/Forskning/Publikationer/Rapporter/2018/papperspedagoger--larares-arbete-med-administration-i-digitaliseringens-tidevarv/
Sandahl, J., Tväråna, M. & Jakobsson, M. (2022). Samhällskunskap (social science education) in Sweden. *Journal of Social Science Education*, *21*(3), 85–106. https://doi.org/10.11576/JSSE-5339
Schüllerqvist, B. & Osbeck, C. (Red.). (2009). *Ämnesdidaktiska insikter och strategier: Berättelser från gymnasielärare i samhällskunskap, geografi, historia och religionskunskap*. Karlstad University Press.
Selghed, B. (2011). *Betygen i skolan: Kunskapssyn, bedömningsprinciper och lärarpraxis* (2 uppl.). Liber.
Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795
Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 226–235. https://doi.org/10.1080/0969594X.2015.1072084
Skolinspektionen. (2010). *Betygssättning i gymnasieskolan* (Skolinspektionens rapport 2010:12). Skolinspektionen. https://www.skolinspektionen.se/globalassets/02-beslut-rapporter-stat/granskningsrapporter/tkg/2010/betygsattning-i-gymnasieskolan/sammanfattning-betyg-gy.pdf
Skolinspektionen. (2021). *Ombedömning av nationella prov 2019*. Skolinspektionen. https://www.skolinspektionen.se/beslut-rapporter-statistik/publikationer/regeringsrapporter/2020/ombedomning-av-nationella-prov-2019/
Skolverket. (2009). *Likvärdig betygssättning i gymnasieskolan? En*

analys av sambandet mellan nationella prov och kursbetyg. (Rapport 2009:338). Skolverket.
Skolverket. (2011a). *Gymnasieskola 2011*. Skolverket.
Skolverket. (2011b). *Planering och genomförande av undervisningen: För grundskolan, grundsärskolan, specialskolan och sameskolan*. Skolverket.
Skolverket. (2012). *Bedömning och betygssättning i gymnasieskolan*. Skolverket.
Skolverket. (2018). *Betyg och betygssättning*. Skolverket.
Skolverket. (2020). *Likvärdiga betyg och meritvärden* (Rapport 2020:7). Skolverket. https://www.skolverket.se/publikationsserier/rapporter/2020/likvardiga-betyg-och-meritvarden
Skolverket. (2021). *Att planera, bedöma och ge återkoppling: Stöd för undervisning*. Skolverket.
Skolverket. (2022a). *Betyg och prövning*. Skolverket.
Skolverket. (2022b). *Läroplan för gymnasieskolan: Reviderad 2022*. Skolverket.
Skolverket. (2022c, 20 december). *Samhällskunskap*. Samhällskunskap. https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/gymnasieprogrammen/amne?url=-996270488%2Fsyllabuscw%2Fjsp%2Fsubject.htm%3FsubjectCode%3DSAM%26version%3D4%26tos%3Dgy&sv.url=12.5dfee44715d35a5cdfa92a3
Skolverket. (2023, 13 april). *Sätta betyg i gymnasieskolan*. https://www.skolverket.se/undervisning/gymnasieskolan/betyg-i-gymnasieskolan/satta-betyg-i-gymnasieskolan
Strandler, O. (2017). *Performativa lärarpraktiker* [Doktorsavhandling, Göteborgs universitet]. Gothenburg studies in educational sciences, 400. https://gupea.ub.gu.se/handle/2077/52325
Sundqvist, P., Sandlund, E., Skar, G. B. & Tengberg, M. (2020). Effects of rater training on the assessment of L2 English oral proficiency. *Nordic Journal of Modern Language Methodology*, *8*(1), 3–29. https://doi.org/10.46364/njmlm.v8i1.605
Svenska Akademien. (2023). Resonera. I *Svenska Akademiens ordbok*. https://www.saob.se/artikel/?unik=R_1263-0172.lP83&pz=3
Svingby, G. (1989). *SO i fokus: Exempel på en undervisning som utgår från begrepp och färdigheter*. Utbildningsförlaget.
Svingby, G. (1998). *Utvärdering av grundskolan 1995: UG 95*. Statens skolverk.
Tengberg, M., Borgström, E., Lötmarker, L., Skar, G. B., Sundqvist, P.,

Walkert, M. & Wikberg, K. (2017). *Likvärdig bedömning av elevers språkförmågor. Preliminära resultat från ett ämnesdidaktiskt forskningsprojekt*. Karlstads universitet.
Teorell, J. & Svensson, T. (2007). *Att fråga och att svara: Samhällsvetenskaplig metod*. Liber.
Tholin, J. (2003). *En roliger dans? Svenska skolors första tolkning av innebörden i lokala betygskriterier i tre ämnen för skolår åtta* [Licentiatuppsats, Högskolan i Borås]. Institutionen för pedagogik. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A876778&dswid=5652
Tierney, R. D. (2013). Fairness in Classroom Assessment. I J. H. McMillan (Red.), *SAGE Handbook of research on classroom assessment* (s. 124–144). SAGE Publications. https://doi.org/10.4135/9781452218649.n8
Torrez, C. A. & Claunch-Lebsack, E. A. (2014). The present absence: Assessment in social studies classrooms. *Action in Teacher Education*, *36*(5–6), 559–570. https://doi.org/10.1080/01626620.2014.977756
Trost, J. & Hultåker, O. (2016). *Enkätboken* (5 uppl.). Studentlitteratur.
Tveit, S. (2018). (Trans)national trends and cultures of educational assessment: Reception and resistance of national testing in Sweden and Norway during the twentieth century. I C. Alarcón & M. Lawn (Red.), *Assessment cultures: Historical perspectives* (s. 131–161). Peter Lang.
Törnvall, M. (2001). *Uppfattningar och upplevelser av bedömning i grundskolan* [Licentiatuppsats, Malmö högskola]. Lärarutbildningen. https://mau.diva-portal.org/smash/record.jsf?pid=diva2%3A1404526&dswid=9653
Uljens, M. (Red.). (1997). *Didaktik: Teori, reflektion och praktik*. Studentlitteratur.
Unal, A. & Unal, Z. (2019). An examination of K-12 teachers' assessment beliefs and practices in relation to years of teaching experience. *Georgia Educational Researcher*, *16*(1), 1–21. https://doi.org/10.20429/ger.2019.160102
Vanlommel, K., Van Gasse, R., Vanhoof, J. & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, *83*, 75–83. https://doi.org/10.1016/j.ijer.2017.02.013
Vetenskapsrådet. (2017). *God forskningssed* (Reviderad utgåva). Vetenskapsrådet.
Vogt, B. (2017). *Just assessment in school: A context-sensitive comparative study of pupils' conceptions in Sweden and Germany* [Doktorsavhandling, Linnéuniversitetet]. Linnaeus University

Dissertations, 298. https://lnu.diva-portal.org/smash/record.jsf?pid=diva2%3A1160059&dswid=-6786

Wahlström, N., Adolfsson, C.-H. & Vogt, B. (2020). Making social studies in standards-based curricula. *JSSE - Journal of Social Science Education*, *19*(SI), 66–81. https://doi.org/10.4119/JSSE-2361

Wahlström, N. & Sundberg, D. (2015). *Theory-based evaluation of the curriculum Lgr 11* [Working paper]. IFAU - Institutet för arbetsmarknads- och utbildningspolitisk utvärdering. https://www.ifau.se/en/Research/Publications/Working-papers/2015/Theory-based-evaluation-of-the-curriculum-Lgr-11/

Wellberg, S. (2023). Teacher-made tests: Why they matter and a framework for analysing mathematics exams. *Assessment in Education: Principles, Policy & Practice*, *30*(1), 53–75. https://doi.org/10.1080/0969594X.2023.2189565

Wermke, W. & Salokangas, M. (2021). *The autonomy paradox: Teachers' perceptions of self-governance across Europe*. Springer International Publishing.

Wherfel, Q. M., Monda-Amaya, L. & Shriner, J. G. (2022). General education teacher practices: Assessment, decision-making and the influence of co-teaching. *Preventing School Failure: Alternative Education for Children and Youth*, *66*(1), 42–51. https://doi.org/10.1080/1045988X.2021.1934650

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, *37*(1), 5–20. https://doi.org/10.1111/emip.12188

Zhang, Z. & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, *16*(4), 323–342. https://doi.org/10.1207/S15324818AME1604_4

Öberg, J. & Bäckström, P., 1983. (2021). Kanon eller inte kanon? En enkätstudie bland samhällskunskapslärare i årskurs 7-9. *Nordidactica. Journal of Humanities and Social Science Education*, *11*(2), 20–46.

Bilagor

Intervjuguider finns som bilagor i delstudie I respektive III.

Bilaga 1 – Kursplan Samhällskunskap 1a1 (2019)

Centralt innehåll

Undervisningen i kursen ska behandla följande centrala innehåll:

- Demokrati och politiska system på lokal och nationell nivå samt inom EU. Internationella och nordiska samarbeten. Medborgarnas möjligheter att påverka politiska beslut på de olika nivåerna. Maktfördelning och påverkansmöjligheter i olika system och på olika nivåer utifrån olika demokratimodeller och den digitala teknikens möjligheter. Digitalisering och mediers innehåll samt nyhetsvärdering när det gäller frågor om demokrati och politik.
- De mänskliga rättigheterna; vilka de är, hur de förhåller sig till stat och individ och hur man kan utkräva sina individuella och kollektiva mänskliga rättigheter.
- Folkrätten i väpnade konflikter. Den internationella humanitära rätten och skyddet för civila i väpnade konflikter.
- Arbetsmarknad, arbetsrätt och arbetsmiljö. Arbetsmarknadens parter, deras olika roller och betydelse för samhällsutvecklingen. Digitaliseringens påverkan på arbetsmarknaden.
- Gruppers och individers identitet, relationer och sociala livsvillkor med utgångspunkt i att människor grupperas utifrån kategorier som skapar både gemenskap och utanförskap.
- Privatekonomi. Hushållets inkomster, utgifter, tillgångar och skulder. Konsumenträtt samt konsumtion i förhållande till behov och resurser. Hur privatekonomin påverkas av samhällsekonomiska förändringar.
- Källkritik. Metoder för att söka, kritiskt granska, värdera och bearbeta information från källor i digital och annan form.
- Presentation i olika former, till exempel debatter och debattinlägg.

Kunskapskrav

Betyget E

Eleven kan **översiktligt** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna. Eleven kan också **översiktligt** redogöra för de mänskliga rättigheterna. I sin analys förklarar eleven **enkla** samband och drar **enkla** slutsatser om likheter och skillnader mellan olika samhällens organisation. Dessutom kan eleven dra **enkla** slutsatser om hur nutida samhällsförhållanden påverkar och påverkas av individer, grupper och samhällsstrukturer.

Eleven kan analysera samhällsfrågor och identifiera **någon** orsak och konsekvens. I analysen diskuterar eleven **översiktligt** orsakerna och konsekvenserna samt möjliga lösningar på samhällsfrågorna. Eleven kan ge **enkla** argument för sina ståndpunkter och värderar med **enkla** omdömen andras ståndpunkter.

Eleven kan **översiktligt** redogöra för individens rättigheter och skyldigheter i rollen som konsument, förhållandet mellan hushållets inkomster och utgifter, tillgångar och skulder samt för sambanden mellan den privata ekonomin och samhällsekonomin.

I arbetet med samhällsfrågor kan eleven **med viss säkerhet** söka, granska och tolka information från olika källor samt göra **enkla** reflektioner om källornas relevans och trovärdighet.

Eleven kan, **med viss säkerhet** och på ett strukturerat sätt, uttrycka sina kunskaper i samhällskunskap i olika presentationsformer.

Betyget D
Betyget D innebär att kunskapskraven för E och till övervägande del för C är uppfyllda.

Betyget C
Eleven kan **utförligt** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna. Eleven kan också **utförligt** redogöra för de mänskliga rättigheterna. I sin analys förklarar eleven samband och drar **välgrundade** slutsatser om likheter och skillnader mellan olika samhällens organisation. Dessutom kan eleven dra **välgrundade** slutsatser om hur nutida samhällsförhållanden påverkar och påverkas av individer, grupper och samhällsstrukturer.

Eleven kan analysera samhällsfrågor och identifierar **några** orsaker och konsekvenser. I analysen diskuterar eleven **utförligt** orsakerna och konsekvenserna samt möjliga lösningar på samhällsfrågorna. Eleven kan ge **välgrundade** argument för sina ståndpunkter och värderar med **enkla** omdömen andras ståndpunkter.

Eleven kan **utförligt** redogöra för individens rättigheter och skyldigheter i rollen som konsument, förhållandet mellan hushållets inkomster och utgifter, tillgångar och skulder samt för sambanden mellan den privata ekonomin och samhällsekonomin.

I arbetet med samhällsfrågor kan eleven **med viss säkerhet** söka, granska och tolka information från olika källor samt gör **välgrundade** reflektioner om källornas relevans och trovärdighet utifrån syftet. Eleven kan, **med viss säkerhet** och på ett strukturerat sätt, uttrycka sina kunskaper i samhällskunskap i olika presentationsformer **samt formulera sig självständigt i förhållande till källorna.**

Betyget B
Betyget B innebär att kunskapskraven för C och till övervägande del för A är uppfyllda.

Betyget A
Eleven kan **utförligt och nyanserat** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna. Eleven kan också **utförligt och nyanserat** redogöra för de mänskliga rättigheterna. I sin analys förklarar eleven **komplexa** samband och drar **välgrundade och nyanserade** slutsatser om likheter och skillnader mellan olika samhällens organisation. Dessutom kan eleven dra **välgrundade och nyanserade** slutsatser om hur nutida samhällsförhållanden påverkar och påverkas av individer, grupper och samhällsstrukturer.

Eleven kan analysera samhällsfrågor och identifierar **flera** orsaker och konsekvenser. I analysen diskuterar eleven **utförligt och nyanserat** orsakerna och konsekvenserna samt möjliga lösningar på samhällsfrågorna. Eleven kan ge **välgrundade och nyanserade** argument för sina ståndpunkter och värderar med **nyanserade** omdömen andras ståndpunkter.

Eleven kan **utförligt och nyanserat** redogöra för individens rättigheter och skyldigheter i rollen som konsument, förhållandet mellan hushållets inkomster och utgifter, tillgångar och skulder samt för sambanden mellan den privata ekonomin och samhällsekonomin.

I arbetet med samhällsfrågor kan eleven **med säkerhet** söka, granska och tolka information från olika källor samt göra **välgrundade och nyanserade** reflektioner om källornas relevans och trovärdighet utifrån syftet. Eleven kan, **med säkerhet** och på ett strukturerat sätt, uttrycka sina kunskaper i samhällskunskap i olika presentationsformer **samt formulera sig självständigt i förhållande till källorna.**

Bilaga 2 – Kursplan i samhällskunskap 1b (2019)

Centralt innehåll

Undervisningen i kursen ska behandla följande centrala innehåll:

- Demokrati och politiska system på lokal och nationell nivå samt inom EU. Internationella och nordiska samarbeten. Medborgarnas möjligheter att påverka politiska beslut på de olika nivåerna. Maktfördelning och påverkansmöjligheter i olika system och på olika nivåer utifrån grundläggande demokratimodeller. Möjligheter och utmaningar med digitaliseringen när de gäller frågor om demokrati och politik. Politiska ideologier och deras koppling till samhällsbyggande och välfärdsteorier.
- De mänskliga rättigheterna; vilka de är, hur de förhåller sig till stat och individ och hur man kan utkräva sina individuella och kollektiva mänskliga rättigheter.
- Folkrätten i väpnade konflikter. Den internationella humanitära rätten och skyddet för civila i väpnade konflikter.
- Arbetsmarknad, arbetsrätt och arbetsmiljö. Arbetsmarknadens parter, deras olika roller och betydelse för samhällsutvecklingen. Digitaliseringens påverkan på arbetsmarknaden.
- Gruppers och individers identitet, relationer och sociala livsvillkor med utgångspunkt i att människor grupperas utifrån kategorier som skapar både gemenskap och utanförskap.
- Samhällsekonomi, till exempel ekonomiska strukturer och flöden i Sverige och internationellt. Försörjning, tillväxt och företagande, resursanvändning och resursfördelning utifrån olika förutsättningar.
- Privatekonomi. Hushållets inkomster, utgifter, tillgångar och skulder. Konsumenträtt samt konsumtion i förhållande till behov och resurser. Hur privatekonomin påverkas av samhällsekonomiska förändringar.
- Mediers och informationsteknikens roll i samhället. Deras möjligheter att påverka människor och samhällsutvecklingen samt de möjligheter de ger människor att påverka. Mediers innehåll och nyhetsvärdering.
- Källkritik. Metoder för att söka, kritiskt granska, värdera och bearbeta information från källor i digital och annan form.
- Samhällsvetenskapliga begrepp, teorier, modeller och metoder i samband med undersökningar av samhällsfrågor och samhällsförhållanden. Exempel på metoder för att samla in information är intervju, enkät och observation. Exempel på metoder för att bearbeta information är statistiska metoder, samhällsvetenskaplig textanalys, argumentationsanalys.
- Presentation i olika former, till exempel debatter, debattinlägg och rapporter.

Kunskapskrav

Betyget E

Eleven kan **översiktligt** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna. Eleven kan också **översiktligt** redogöra för de mänskliga rättigheterna. I sin analys förklarar eleven **enkla** samband och drar **enkla** slutsatser om likheter och skillnader mellan olika samhällens organisation. Dessutom kan eleven **översiktligt** redogöra för de historiska förutsättningarnas betydelse och dra **enkla** slutsatser om hur nutida samhällsförhållanden påverkar och påverkas av individer, grupper och samhällsstrukturer.

Eleven kan analysera samhällsfrågor och identifiera **någon** orsak och konsekvens. I analysen använder eleven **med viss säkerhet** samhällsvetenskapliga begrepp, teorier, modeller och metoder. Eleven diskuterar **översiktligt** orsakerna och konsekvenserna samt möjliga lösningar på samhällsfrågorna. Eleven kan ge **enkla** argument för sina ståndpunkter och värderar med **enkla** omdömen andras ståndpunkter.

Eleven kan **översiktligt** redogöra för individens rättigheter och skyldigheter i rollen som konsument, förhållandet mellan hushållets inkomster och utgifter, tillgångar och skulder samt för sambanden mellan den privata ekonomin och samhällsekonomin.

I arbetet med samhällsfrågor kan eleven **med viss säkerhet** söka, granska och tolka information från olika källor, redovisa sina källor samt göra **enkla** reflektioner om deras relevans och trovärdighet. Eleven kan, **med viss säkerhet** och på ett strukturerat sätt, uttrycka sina kunskaper i samhällskunskap i olika presentationsformer.

Betyget D innebär att kunskapskraven för E och till övervägande del för C är uppfyllda.

Betyget C

Eleven kan **utförligt** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna. Eleven kan också **utförligt** redogöra för de mänskliga rättigheterna. I sin analys förklarar eleven samband och drar **välgrundade** slutsatser om likheter och skillnader mellan olika samhällens organisation. Dessutom kan eleven **utförligt** redogöra för de historiska förutsättningarnas betydelse och dra **välgrundade** slutsatser om hur nutida samhällsförhållanden påverkar och påverkas av individer, grupper och samhällsstrukturer.

Eleven kan analysera samhällsfrågor och identifiera **några** orsaker och konsekvenser. I analysen använder eleven **med viss säkerhet** samhällsvetenskapliga begrepp, teorier, modeller och metoder **samt värderar dem med enkla omdömen**. Eleven diskuterar **utförligt** orsakerna och konsekvenserna samt möjliga lösningar på samhällsfrågorna. Eleven kan ge **välgrundade** argument för sina ståndpunkter och värderar med **enkla** omdömen andras ståndpunkter.

Eleven kan **utförligt** redogöra för individens rättigheter och skyldigheter i rollen som konsument, förhållandet mellan hushållets inkomster och utgifter, tillgångar och skulder samt för sambanden mellan den privata ekonomin och samhällsekonomin.

I arbetet med samhällsfrågor kan eleven **med viss säkerhet** söka, granska och tolka information från olika källor, redovisa sina källor samt göra **välgrundade** reflektioner om deras relevans och trovärdighet utifrån syftet. Eleven kan, **med viss säkerhet** och på ett strukturerat sätt, uttrycka sina kunskaper i samhällskunskap i olika presentationsformer **samt formulera sig självständigt i förhållande till källorna.**

Betyget B innebär att kunskapskraven för C och till övervägande del för A är uppfyllda.

Betyget A

Eleven kan **utförligt och nyanserat** redogöra för och analysera olika samhällens organisation och samhällsförhållanden samt de bakomliggande idéerna. Eleven kan också **utförligt och nyanserat** redogöra för de mänskliga rättigheterna. I sin analys förklarar eleven **komplexa** samband och drar **välgrundade** slutsatser om likheter och skillnader mellan olika samhällens organisation. Dessutom kan eleven **utförligt och nyanserat** redogöra för de historiska förutsättningarnas betydelse och dra **välgrundade och nyanserade** slutsatser om hur nutida samhällsförhållanden påverkar och påverkas av individer, grupper och samhällsstrukturer.

Eleven kan analysera samhällsfrågor och identifierar **flera** orsaker och konsekvenser. I analysen använder eleven **med säkerhet** samhällsvetenskapliga begrepp, teorier, modeller och metoder **samt värderar dem med nyanserade omdömen**. Eleven diskuterar **utförligt och nyanserat** orsakerna och konsekvenserna samt möjliga lösningar på samhällsfrågorna. Eleven kan ge **nyanserade** argument för sina ståndpunkter och värderar med **nyanserade** omdömen andras ståndpunkter.

Eleven kan **utförligt och nyanserat** redogöra för individens rättigheter och skyldigheter i rollen som konsument, förhållandet mellan hushållets inkomster och utgifter, tillgångar och skulder samt för sambanden mellan den privata ekonomin och samhällsekonomin.

I arbetet med samhällsfrågor kan eleven **med säkerhet** söka, granska och tolka information från olika källor, redovisa sina källor samt göra **välgrundade och nyanserade** reflektioner om deras relevans och trovärdighet utifrån syftet. Eleven kan, **med säkerhet** och på ett strukturerat sätt, uttrycka sina kunskaper i samhällskunskap i olika presentationsformer **samt formulera sig självständigt i förhållande till källorna.**

Bilaga 3 – Missivbrev för den första datainsamlingen

Hej!
Jag heter Tobias Jansson och går en forskarutbildning inom lärarlyftet. Min forskning handlar om skriftliga prov i samhällskunskap; vilka olika former av kunskap testas i prov och vilka faktorer påverkar utformningen av skriftliga prov. Det finns väldigt lite kunskap om bedömning i skolan. Din medverkan skulle därför bidra till en bättre förståelse av lärares praxis och kanske också bidra till en skolutveckling. Jag behöver din hjälp med två saker.

1. Min önskan är att du skickar två ***skriftliga prov*** du använde under en kurs i Samhällskunskap A som avslutades läsåret 2008-2009 i *en studieförberedande klass* och/eller *en yrkesförberedande klass*. (Om du bara undervisade på den ena typen av program så skicka bara proven för dessa.)

   Proven ska behandla arbetsområdena *samhällsekonomi* respektive *Sveriges styrelseskick*. (Om dina arbetsområden har en annan indelning ber jag dig skicka in de prov som närmast täcker in de områden jag angett). Det innebär att du skickar in *maximalt fyra* prov. Fyll för varje prov i de bifogade försättsbladen.

2. Vidare önskar jag att du fyller i den medföljande ***enkäten***. Naturligtvis kommer alla uppgifter att hanteras konfidentiellt och alla deltagare garanteras anonymitet.

Det kan bli aktuellt med några uppföljande intervjuer av några lärare, vilka väljs utifrån deras prov. Därför behöver jag kunna se vem som gjort respektive prov. Det är dock naturligtvis frivilligt att uppge namn. Ingen annan kommer att ha tillgång till dessa uppgifter och när studien är klar kommer dessa uppgifter att raderas. I avhandlingen kommer endast fingerade namn på personer och skolor att användas. Det går bra att senare välja att inte delta i en intervju, även om du samtycker nu.

Om du samtycker till att kunna kontaktas för en eventuell senare intervju undertecknar du ***blanketten om samtycke***, annars är det bara att hoppa över det. Om du inte vill ställa upp på en senare intervju, är jag tacksam om du ändå skickar in enkät och prov anonymt.

Skicka ***prov*** (med försättsblad), ***enkät*** och ***blanketten om samtycke*** i det bifogade svarskuvertet. Koden på svarskuvertet är endast till för att veta vilka som besvarat enkäten och registreras inte. Svara senast måndagen den **12 oktober 2009**.

Tack för din medverkan!
Tobias Jansson, forskarstuderande

Blankett om samtycke

Jag samtycker till att ställa upp på en eventuell senare intervju och låter mig identifieras under forskningsprocessen. Jag är informerad om att jag när som helst kan avstå från intervjun och att jag även senare kan välja att inte medverka i studien.

…….
(Underskrift)

…….
(Namnförtydligande)

Arbetar på skolan:

…….
(Skolans namn)

(Detta blad kommer att rivas bort och förstöras efter studiens avslutande så att ingen kan identifiera dem som deltagit i studien.) [8]

[8] Missivbrevet för intervjuerna skickades via en e-postadress som jag inte längre har tillgång till och tyvärr finns ingen kopia sparad av detta.

Bilaga 4 – Missivbrev för den andra datainsamlingen

Forskning om bedömning

Jag heter Tobias Jansson och arbetar som lärare vid Linköpings universitet och ska genomföra ett forskningsprojekt kring bedömning i samhällskunskap. Jag undrar om du har möjlighet att bidra med material till min forskning. Min undersökning handlar om att studera de underlag som samhällskunskapslärare på gymnasiet använder för bedömning av elevernas kunskapsutveckling. Fokus i studien är dels i vilka former elevernas kunskapsutveckling bedöms, dels vilka kunskapsformer som bedöms i de olika formerna. Jag har avgränsat studien till kurserna Samhällskunskap 1a1 och 1b.

Det material jag således önskar ta del av är en så komplett uppsättning som möjligt av de bedömningsunderlag som du använt i kurserna Samhällskunskap 1a1 respektive 1b (markera vilket material som hör till vilken kurs). Om du bara haft en av kurserna så räcker det naturligtvis med den. Det som avses med underlag är allt det som du använt för att bedöma elevernas kunskapsutveckling. Det kan handla om prov, inlämningsuppgifter, muntliga redovisningar, uppsatser/rapporter, rollspel etc. Dessa kan ha använts för såväl summativa (betygsgrundande) som formativa (för lärande) syften. Det jag önskar ta del av är själva uppgiftsbeskrivningen, inte elevernas svar. Om det finns någon form av bedömningskriterier (t.ex. en matris) till uppgiften önskar jag även ta del av detta.

Jag är medveten om att det kan ta lite tid att sammanställa detta, men är oerhört tacksam om du har möjlighet att göra detta och bidra till forskningen och en bättre förståelse av hur lärares bedömningar går till. Förhoppningsvis kan detta också bidra till att ge oss kunskap till att ytterligare utveckla lärares arbete med bedömningar.

Enklast är om du kan maila materialet till mig, annars kan jag skicka ett svarskuvert om det behövs. Jag behöver veta vem materialet kommer ifrån för att kunna kontakta dig om jag har några frågor. Det är dock bara jag som kommer att ha tillgång till denna information. Under arbetsprocessen och i avhandlingen kommer allt material att anonymiseras.

Om du har några frågor är du välkommen att kontakta mig.

Tack för din medverkan!
Tobias Jansson, universitetsadjunkt

Tobias Jansson

Hej NN,

Tack än en gång för allt material som du skickade till mig i våras. Dina bedömningsuppgifter är mycket intressanta och bidrar till en bättre bild av lärares bedömningspraktik. I analysen av dessa har jag nu insett att det vore bra att få veta mer om sammanhanget och om dina tankar kring dina bedömningsuppgifter. Jag skulle därför väldigt gärna vilja träffa dig för en intervju kring din bedömningspraktik.

Jag vet att du redan avsatt en del av din värdefulla tid genom att ta fram och skicka bedömningsuppgifterna till mig. Jag hoppas att du ändå även skulle kunna avsätta ytterligare max 1½ timme för en intervju. Jag tänker mig att jag i så fall kommer till din arbetsplats (om du föredrar en annan plats så går naturligtvis det också bra) vid ett tillfälle som passar dig i slutet av januari eller i februari.

Intervjun är självklart helt frivillig, du kan välja att avbryta den när du vill eller senare välja att inte medverka i studien. Dina svar kommer också att anonymiseras.

Med förhoppning om din medverkan!

Vänliga hälsningar
/ Tobias Jansson

Delstudier

De till denna avhandling tillhörande delstudier har tagits bort av upphovsrättsliga skäl. För mer information gå till:

https://doi.org/10.3384/9789180754477

UTBILDNINGSVETENSKAP

Linköping Studies in Behavioural Science No. 256, 2024
Institutionen för beteendevetenskap och lärande

Linköpings universitet
581 83 Linköping

www.liu.se

LINKÖPINGS UNIVERSITET