

Beställningsadress:
fritzes kundservice 106 47 stockholm

tel: 08-690 95 76 fax: 08-690 95 50

e-postadress: skolutveckling@fritzes.se
www.skolutveckling.se

Best.nr. U03:010

helena korp

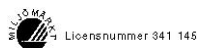
Kunskapsbedömning
– *hur, vad och varför*

ISBN 91-85128-05-8
ISSN 1651-3460

Omslagsbild: pierre rosberg, tiofoto ab

Tryck: lenanders grafiska ab, kalmar 2003 · 12502

Svanenmärkt trycksak MT



Licensnummer 341 145

Kunskapsbedömning

– hur, vad och varför

HELENA KORP

MYNDIGHETEN FÖR SKOLUTVECKLING

innehållsförteckning	
förord.	9
sammanfattning.	11
summary.	13
inledning.	15
Bakgrund.	16
Syfte.	18
Uppläggnig.	19
Metod.	20
inledning.	15
Bakgrund.	16
Syfte.	18
Uppläggnig.	19
Metod.	20
kunskapsbedömning i ett historiskt perspektiv.	25
de första institutionella bedömningarna.	26
intelligenstestningens tidsålder.	31
efterkrigstidens utbildningsexpansion.	37
Objektivt, rättvist, tillförlitligt.	38
1970-talets utbildningsreformer.	41
bedömning i det postmoderna samhället.	43
Motivera istället för att sortera.	44
Ny kunskapssyn.	45
Den sociala ordningen i svenska klassrum.	48
Bedömningens konsekvenser.	51
Internationella tendenser.	53
Vad innebär de ”det nya provparadigmet”?	55
begreppsliga ramar.	58
”det var enklare för”	58
Teoretisk omorientering.	58
inlärningsteorier.	62
Behaviourismen.	63
Kognitiva och metakognitiva teorier.	64
Sociokulturella perspektiv.	67

hur kan man kategorisera prov?	70
bedömning på individ- och systemnivå.	72
Bedömning av individers kunskaper.	74
Bedömningens betydelse	76
Bedömningens syfte	77
<i>Summativ bedömning</i>	77
<i>Certifikation</i>	78
<i>Differentiering</i>	78
<i>Kompetens och kvalifikation</i>	78
<i>Kvalitetssäkring samhällsekonomiska intressen</i>	79
<i>Formativ bedömning av enskilda elever</i>	79
<i>Formativ bedömning och undervisning</i>	82
<i>Formativ bedömning i mångkulturell utbildning</i>	83
vad är det som skall bedömas?	85
Kunskap	85
<i>Kunskapsbegreppet i LPO 94</i>	85
<i>Taxonomier</i>	87
<i>Kompetens</i>	90
<i>Behållning och prestation</i>	91
<i>Förmåga</i>	92
olika provformer	93
”Konventionella” och ”alternativa” prov	93
Objektiva prov	96
Långa och korta svar	97
Autentiska prov	98
Portfolio	99
När passar vilket prov?	100
att värdera elevers prestationer	103
Kriterierelaterad bedömning	103
Normrelaterad bedömning	104
Att bedöma sammansatta kunskaper med god tillförlitlighet	107
Global och analytisk bedömning	112
Den kriterierelaterade bedömningens kunskapsteori	112
Naturvetenskapligt orienterad bedömning	113

Bedömningsmodeller	113
Humanistiskt eller konstnärligt orienterad bedömning .	114
Hur kan resultaten rapporteras?	116
bedömningens kvalitet	118
”Mätmodellen” och ”standardmodellen”	118
Kvalitetsbegrepp	119
<i>Reliabilitet</i>	120
<i>Validitet</i>	120
<i>Relationen mellan reliabilitet och validitet i bedömningen .</i>	123
Alternativa sätt att resonera om kvalitet i bedömning . .	125
faktorer som kan störa bedömnings-	
resultatet	126
Faktorer som påverkar prestationen	126
” <i>Test-smartness</i> ”	126
<i>Motivation och ängslan</i>	127
<i>Språkliga hinder</i>	128
<i>Artificiella bedömningssituationer</i>	129
Faktorer som påverkar bedömningsprocessen	130
<i>Kulturella faktorer</i>	131
Objektivitet och rättvisa i bedömningsprocessen	133
slutord	135
Ideologiska perspektiv på ”det nya provparadigmet” . . .	136
Prov i praktiken – konsekvenser	142
Frågor för fortsatt forskning	146
referenser	149
bilaga 1	163
serien forskning i fokus vid myndigheten för skolutveckling ..	171
Tidigare forsknings- och kunskapsöversikter som publicerats	171

Förord

I denna bok presenterar Helena Korp en översikt av forskningen inom området kunskapsbedömning. Vetenskaplig ledare har professor Gunilla Svingby varit. Det finns behov av en sådan sammanställning bland forskare. Syftet är att i denna skrift kombinera detta med behovet bland lärare, verksamma på olika nivåer i utbildningssystemet, att reflektera över och utveckla metoder för att bedöma elevers kunskaper.

Hur kan man veta vad någon annan kan och vet? Vilka kunskaper är värda att överföra till nästa generation? Hur lär sig elever? Vilken roll ska skolan ha i samhället? Vad har elever rätt att lära sig i skolan? Är det nödvändigt att jämföra elever? Kan sådana jämförelser någonsin bli meningsfulla och rättvisa? Detta är några av de frågor som författaren tar upp till diskussion.

Författaren gör en historisk tillbakablick som relateras till olika utbildningssystemens expansion samt politiska och ekonomiska förändringar. Hon gör ett försök att beskriva men också problematisera vad som ibland kallas ”det nya provparadigmet”. Här tecknas också ett begreppsligt ramverk för diskussioner om kunskapsbedömning. I framställningen beskrivs hur olika typer av bedömning används på olika nivåer i utbildningssystemet och vilka syften som kan finnas med bedömning på de olika nivåerna. Huvudfokus ligger på sådana bedömningar som lärare gör av sina elever i klassrummet. Arbetet kan delvis ses som en form av diskursanalys, eftersom en viktig del i det har varit att ”dekonstruera” vetenskapliga texter om bedömning, för att få syn på underliggande ideologiska och teoretiska mönster. Ett viktigt mål är alltså att synliggöra och begripliggöra dessa mönster för läsaren.

Det finns ett stort internationellt intresse för bedömningsforskning. Frågor om bedömning har också haft en framträdande roll på internationella forskarkonferenser. Däremot har det vetenskapliga intresset varit tämligen svalt i Sverige och i Norden. Handböcker, läroböcker mm som sätter in prov och

bedömning i teoretiskt, historiskt och samhälleligt sammanhang har märkligt nog i stort sett saknats på svenska. Likaså är de empiriska studierna inom området förhållandevis få i Sverige. Ofta har de haft karaktären av utvärderingar. Detta är anledningen till att Skolverket numera Myndigheten för skolutveckling för några år sedan valde att stödja uppbyggnaden av ett nätverk inom bedömningsområdet och arrangerade en konferens för att diskutera olika forskningsbehov. Inom nätverket har diskussioner och aktiviteter stimulerats. Initiativet till föreliggande översikt togs genom nätverkets arbete. Förhoppningsvis bidrar denna publikation till att ytterligare aktivera forskare och stimulera till samtal och aktiviteter inom praktisk pedagogisk verksamhet.

Annika Andræ Thelin

FORSKNINGSCHEF

Sammanfattning

I boken diskuteras vad kunskapsbedömning används till, hur den sker, vilka slags kunskaper som olika prov och andra slags bedömningar syftar till att "komma åt" och vilka konsekvenser valet av bedömningsform får för lärandet, men också för de studerandes rättssäkerhet och för utbildningen som sådan. Såväl pedagogiska som sociologiska och kunskapsteoretiska aspekter tas upp.

Studien bygger på en genomgång av forskningslitteratur av olika slag, främst från den anglosaxiska delen av världen, där forskningen om kunskapsbedömning har varit mycket omfattande sedan mitten av 1980-talet. Genomgången visar att stora förändringar har skett i den västliga världen under de senaste par decennierna när det gäller hur man ser på och använder bedömning. Flera författare talar om "ett nytt provparadigm". Det är dock inte fråga om något enhetligt synsätt på en djupare nivå, utan olika forskare och aktörer förespråkar en förnyelse av kunskapsbedömningen på olika nivåer i utbildningssystemet utifrån tämligen skilda ideologiska utgångspunkter.

En utgångspunkt för "det nya provparadigmet" är att syftet med kunskapsbedömning har vidgats från betygssättning och sortering till att också stödja individers lärande och utgöra instrument för utveckling av undervisningen i klassrummet och av läroplanerna. Denna syftesförskjutning ska bl. a. förstås mot bakgrund av att stora politiska satsningar har gjorts över hela västvärlden för att få en ökande andel ungdomar att stanna länge i utbildningssystemet och för att få medborgarna att engagera sig i ett "livslångt lärande". Kunskapstillväxten i vetenskapssamhället, informationsteknologins spridning och ökade sociala och kulturella spänningar i samhället har också medfört att stoffvalet har blivit ett allt svårare problem för skolan. Att lära de studerande "att lära" har därför i stor utsträckning, åtminstone i retoriken om utbildning, kommit att ersätta specifika kunskapsmål.

Denna förskjutning av utbildningens mål har för bedömningsforskningen inneburit en teoretisk omorientering, där den kognitiva psykologin och senare också sociokulturella teorier till stora delar har konkurrerat ut behaviourismen som inlärningsteoretisk grund. Bedömningen av de enskilda elevernas kunskaper

per kom i och med detta perspektivskifte att betraktas som ett potentiellt medel för att hjälpa eleverna att förbättra sina lärstrategier. Bedömningen måste därför fokusera lärprocesserna lika väl som resultatet av lärandet och integreras i själva undervisningen. Detta synsätt ställer nya krav på formerna för kunskapsbedömning och på elevernas delaktighet i bedömningsprocessen. Om bedömningen ses som en integrerad del av undervisningen, måste den inte bara mäta, utan också stimulera det lärande som undervisningen syftar till. Såväl svenska som utländska studier har emellertid visat att de mest förekommande proven snarare motverkade än främjade ett förståelseinriktat och kritiskt lärande, framför allt eftersom de fokuserade återgivande av kontextlös kunskap och inte fordrade analys och självständigt tänkande.

De studier av kunskapsbedömning som har genomförts med utgångspunkt i sociokulturell teoribildning betonar lärandets situerade och sociala karaktär. Elevernas förståelse av uppgiften och det sammanhang som den ingår i har då visat sig ha avgörande betydelse för hur de hanterar den. Exempelvis klarar elever ofta av saker i verkliga livet som de inte klarar av i skolan och såväl unga som vuxenstuderande har ofta befunnits oförmögna att tillämpa kunskaper som de har tillägnat sig genom utbildning i verkliga problemsituationer. På basis av sådana resultat har olika slag av "autentisk bedömning" utvecklats. Vygotsky inspirerat en stor mängd studier av bedömning som kommunikativ situation där läraren med hjälp av rätt avpassade uppgifter och adekvat återkoppling hjälper eleven att flytta fram gränserna för sin egen förmåga. Bland dessa återfinns forskning som motiveras av en politisk jämlikhetssträvan och en tro på att en hög kunskapsnivå kan uppnås av alla – inte bara en liten elit.

Det ligger en makt i bedömningsakten – makt över människors identitet, självkänsla och framtidsutsikter, men också över vilka perspektiv och erfarenheter som ska värderas i samhället. Rättvisa och rättssäkerhet är oavvisliga krav på bedömning som påverkar människors livschanser – men kan aldrig fullt ut uppnås. I boken problematiseras tanken om att någon särskild bedömningsmodell skulle kunna utgöra en patentlösning som eliminerar de politiska och etiska dilemman som finns inbyggd i all pedagogisk verksamhet – och i synnerhet i bedömningen av andras kunskaper.

Summary

This book discusses the different uses and methods of educational assessment and what kinds of knowledge those different modes of assessment are designed to evaluate. It also examines how different ways of assessing knowledge affect students' learning and integrity and the educational process. Sociological and epistemological issues are addressed as well as issues of validity.

The study is based on a review of research literature of different kinds, primarily from the Anglo-Saxon parts of the world, where educational assessment has been an expansive research field since the mid-nineteen eighties. The review shows that there have been considerable changes in the Western world in the discourse and practise of educational assessment in the past two decades. Several authors talk about the emergence of a "new assessment paradigm" or a changing assesment culture. However, this concept does not represent a consensus on a deeper level – in fact, researchers and other actors on the educational arena are often proposing similar changes in the practises of educational assessment with quite different ideological motives.

A fundament for "the new assessment paradigm" is that the purposes for educational assessment have been extended from grading and ranking to also encompassing the support of individuals' learning as well as serving the development of classroom-practise and formal curriculum. This extension or shift of purpose should be understood in relation to the fact that vast political efforts have been made in all Western countries to keep an increasing proportion of the young generation in the educational system for a long time and to involve citizens in "life-long-learning". Furthermore, the growth of scientific knowledge, the spread of information technology and increased social and cultural tensions in the society has made the process of content selection an increasingly difficult problem for the school. Teaching the students "to learn" has hence to a large extent, at least on a rhetorical level, come to replace learning-goals that relate to specific subject matter.

To the research on educational assessment, this shifting of learning goals has entailed a theoretical re-orientation, where cognitive psychology and, later, socio-cultural theories to a large extent have come to replace behaviourism as the basis for understanding lear-

ning. As a consequence of this change in perspective, the assessment of individual students' knowledge came to be regarded as a potential mean for helping the students to improve their own learning strategies. The assessment, hence, would have to focus on the learning process as well as the outcomes of learning and be made part of the educational enterprise. This perspective put new demands on the modes of assessment and on the participation of the learners in the assessment procedure. Furthermore, if the assessment is to be regarded as an integrated part of the learning process, it has to not only measure, but also enhance the intended learning. However, Swedish as well as international research has concluded that tests most of the time rather hinder than promote complex and critical learning in the students, mainly because they are focused on the reproduction of de-contextualized knowledge and do not require e.g. analysis, independent thinking or creativity of the students.

Studies of assessment, which are based on socio-cultural theory, emphasise the situated and social nature of learning. They show that students' perception of the task and its context are significant for the way they understand and handle it. E.g. students often manage to do things in real life which they don't manage in school and young as well as mature students are often unable to use knowledge acquired in formal education in real life-situations. On the basis of such findings models for more authentic testing have been developed. Vygotsky has inspired many researchers to examine assessments as communicative situations where the teacher using tasks just above the competency-level of the individual learner and providing adequate support and feed-back help in the solving-process manage to help the learner to extend the boundaries of his competence and improve his general learning strategies. In this line of research there are studies, which are motivated by a political struggle for equity and a belief that excellence is obtainable for all, not just for a small elite.

Assessment entails power – over people's identity, self-worth and future, but also over which perspectives and experiences that should be valued in the society. Fairness and justice are indismisable obligatory demands on assessments that affect persons' life-chances – but they can never be fully achieved. By pointing to the complexity of educational assessment and relating discourses of assessment to their historical background the author hopes to attract the readers attention to the political context and consequences of educational assessment.

Inledning

Hur kan man veta vad någon annan kan och vet? Vilka kunskaper är värda att överföra till nästa generation? Hur lär sig elever? Vilken roll ska skolan ha i samhället? Vad har elever rätt att lära sig i skolan? Är det nödvändigt att jämföra elever? Kan sådana jämförelser någonsin bli meningsfulla och rättvisa?

Varje beslut som gäller bedömning av elevers kunskaper inbegriper ställningstaganden i teoretiska och normativa frågor av det här slaget. Det innebär dock inte, att lärare och andra som sysslar med kunskapsbedömning i sin vardagliga yrkespraktik, uppfattar eller medvetet hanterar sådana kunskaps- och moralfilosofiska, samhälls-, läroplans- och inlärningsteoretiska frågor. Troligen är de mera dolda än uppenbara både för dem som genomför bedömningarna och för dem som bedöms. Med det menar jag att bedömningsprocessen och dess resultat till vardags uppfattas som en självklar del av livet i skolan för såväl lärare som elever och föräldrar. Att pressen från prov och betyg kan upplevas som frustrerande både av lärare och elever och ha förödande konsekvenser för självkänslan och motivationen hos elever som ofta misslyckas, innebär nämligen inte nödvändigtvis att man problematiserar bedömningen som sådan, eller på vad sätt den bedömning som sker i skolan bidrar till att definiera kunskap och sociala relationer i det vidare samhället. Det omvända perspektivet är förmodligen inte heller särskilt vanligt – att proven betraktas inte bara som nödvändiga för att ge underlag till betygssättning eller för att stödja elevernas lärande, utan som instrument för att upprätthålla den sociala ordningen i klassrummet – och i samhället! – i tider när utbildningens bruksvärde inte självklart legitimerar systemet.

Istället, det är min uppfattning, problematiseras proven oftast inom vissa diskursiva ramar, där vissa antaganden och värderingar om lärande, undervisning, kunskap och samhälle tas för givna. För att få en mera reflekterad förståelse av prov och bedömning är det emellertid nödvändigt att försöka komma bortom förståelsen av dem som i denna mening ”naturlig”. Man

måste då fråga sig hur de begreppsliga och ideologiska sammanhang, vari bedömningar och deras resultat framstår som så självklara, ser ut.

Hur formas då våra uppfattningar om bedömning i skolan? Naturligtvis är det fråga om komplexa processer, där socialisation och specifika erfarenheter av utbildningssystemet spelar in, men även den diskursiva nivån är väsentlig på så sätt att den ger oss kategorier för att tolka våra erfarenheter. Vad forskningen om bedömning handlar om och vilka som är dess teoretiska och ideologiska utgångspunkter är särskilt intressant, eftersom forskningen samtidigt som den är en del av diskursen också tillför nya perspektiv och kategorier.

I den här texten försöker jag förmedla den bild som jag genom min läsning av handböcker, vetenskapliga artiklar och forskningsrapporter har skapat mig av den vetenskapliga diskursen om bedömning. Jag har också försökt att sätta in denna i en historisk och samhällelig kontext. Därmed prövar jag i den här skriften att se proven som del av delvis andra sammanhang än vad som har varit vanligt i svensk forskningslitteratur.

Bakgrund

Den internationella forskningen om kunskapsbedömning har sedan mitten av 1980-talet både expanderat avsevärt och genomgått en teoretisk omorientering – framträdande forskare inom området talar till och med om inträdet av en ”ny era” (Birenbaum, 1996), en ”ny bedömningskultur” (Calfee & Masuda, 1997), ett ”nytt provparadigm” (Broadfoot, 1996) och liknande. Den teoretiska omorienteringen består i grova drag av att

- 1) behaviourismen som inlärningsteoretisk grund till stora delar har övergivits till förmån för kognitiva, metakognitiva och sociokulturella teorier om lärande. Enligt dessa innefattar lärande tolkningsprocesser och konstruktion av mening, och i dessa processer, som alltid sker i sociala sammanhang, engageras den lärandes förförståelse, attityder, erfarenheter och

intressen (Birenbaum, 1996). Lärarens uppgift blir då att tillhandahålla meningsfulla uppgifter som har anknytning till elevernas erfarenheter och som förbättrar deras förståelse och deras inlärningsstrategier. I en provsituation kan det innebära, att eleverna får lösa uppgifter tillsammans, att de uppmanas att pröva och utvärdera olika strategier, och att läraren hjälper eleven att prestera sitt bästa, snarare än undanhålla hjälp för att eleven ska prestera sitt mest typiska (t ex Shepard, 2000).

- 2) perceptionen av intelligens som endimensionell, medfödd, konstant och normalfördelad har ersatts av nyare (konstruktivistiska) uppfattningar av intelligens som sammansatt av flera dimensioner och som resultatet av en kombination av biologiska förutsättningar och erfarenheter. Ett viktigt budskap i dessa är att "man kan lära sig att lära": intelligensen har en metakognitiv dimension och genom att denna utvecklas, naturligt eller genom undervisning, så utvecklas intelligensen (Cizek, 1997). Med tillgång till undervisning av hög kvalitet är det därför möjligt för i princip alla, inte bara en liten elit, att uppnå en hög kunskapsstandard (Shepard, 2000; Wolf & Reardon, 1996).
- 3) för bedömningsforskningen nya discipliner, som epistemologi, etik, kunskapssociologi och utbildningshistoria har fått ökad relevans på bekostnad av exempelvis testteori, inläringsteori och yrkessociologi (Lundgren, 1990). Exempelvis är diskussionerna om validitet i "det nya provparadigmet" snarlika de som Messick (1991) för kring produktionen av vetenskaplig kunskap inom poststrukturalistiskt orienterad forskning.
- 4) Inom den internationella bedömningsforskningen har, enligt t ex den brittiska forskaren Patricia Broadfoot (1996) och amerikanska Lorrie Shepard (2000), intresset under de senaste åren till övervägande del varit riktat mot provens pedagogiska, snarare än deras mättekniska kvaliteter. Broadfoot (ibid., s. 15) noterar sålunda bl a att man nu "betonar annorlunda kvalitetskriterier för prov, framförallt deras effekter på lärandet snarare än mätningen i sig", samt att intresset har vuxit för att samla in och tolka sådana uttryck för elevers lärande som gäller "kompetens" och personliga kvaliteter, som betonas i dagens skola".

Ett uttryck för det stora intresse, som internationellt finns för bedömningsforskning är den stora publikationen av vetenskapliga artiklar och böcker i ämnet. Som exempel gav en sökning på termen "assessment" i den pedagogiska databasen ERIC ca 25 000 träffar för publiceringsåren 1997-2000. Dessa representerar i huvudsak tre slag av vetenskapliga publikationer: 1) handböcker, översiktsböcker och läroböcker om bedömning, 2) utvärderingar och beskrivningar av olika slags utvecklingsprojekt inom bedömningsområdet, 3) empiriska studier av exempelvis hur bedömning av olika slag påverkar lärande och undervisning. Frågor om bedömning har de senaste åren också haft en framskjutna plats på internationella forskarkonferenser.

Om man ser till det antal studier som har publicerats inom området i Sverige och Norden under samma period, så kan man dock konstatera att det vetenskapliga intresset för bedömning har varit tämligen svalt i jämförelse med det internationella (se bilaga 1). Av de vetenskapliga publikationer som finns utgivna i dessa länder utgörs den absoluta merparten av utvärderingar och utvecklingsarbeten (ofta exempel på prov som har använts och rapportering av resultat¹). Handböcker och läroböcker som sätter in olika prov och bedömning i ett teoretiskt, historiskt och samhällsligt sammanhang och som speglar den senaste internationella utvecklingen inom området i stort sett har saknats på svenska och de empiriska studierna är mycket få.

Det finns alltså ett stort behov av såväl av svenska hand- och översiktsböcker som empiriska studier av svensk bedömningspraktik.

Syfte

Syftet med denna skrift är

- att sätta de bedömningsformer och system för bedömning som vi har idag i ett historiskt och samhällsligt sammanhang

¹ I Sverige gäller det främst utvecklingen och resultaten av de nationella proven och av högskoleprovet och andra urvalsprov.

- att ge en bild av olika vetenskapliga diskurser om kunskapsbedömning
- att ge en översikt över begrepp och teorier som är centrala inom bedömningsforskningen
- att presentera och diskutera några av de modeller som används för att pröva och värdera elevers kunskaper

Uppläggning

Det andra till och med det sjätte kapitlet utgörs av en historisk tillbakablick, där synen på kunskapsbedömning främst relateras till de nationella utbildningssystemens expansion och de politiska och ekonomiska förändringarna i ”västvärlden” sedan slutet av andra världskriget. Denna första del avslutas med ett försök att dels beskriva vad som ibland har kallats ”det nya provparadigmet” och som representerar mer eller mindre alternativa synsätt på lärande, kunskap och utbildning, dels ge en bild av de ekonomiska och ideologiska förhållanden som kan antas ha påverkat innehållet i ”det nya provparadigmet”.

Det sjunde till och med det artonde kapitlet utgör ett försök att teckna ett begreppsligt ramverk för diskussioner om kunskapsbedömning. Här beskrivs hur olika typer av bedömning används på olika nivåer i utbildningssystemet och vilka syften som kan finnas med bedömning på de olika nivåerna. Framställningens fokus ligger på sådan bedömning som gäller de enskilda elevernas kunskaper, d v s sådan bedömning som lärare gör av sina elever i klassrummet. Denna kan dock inte alltid behandlas helt skilt från större system. Vilka *slags kunskaper* som läraren väljer att pröva, vilka *metoder* som används för att pröva dem och vilka *principer* som tillämpas när resultatet ska värderas, styrs ju i viss mån av de funktioner som bedömningen fyller i utbildningssystemet som helhet och får också konsekvenser utanför klassrummet. Så utgör t ex lärarens bedömningar grunden för elevernas betyg, som å ena sidan är ett medel för den enskilde i konkurrensen om t ex utbildningsplatser, å andra sidan kan användas vid jämförelser av olika skolor. Både prov som utformats av lärarna själva och nationella prov och utvärderingar

påverkar dessutom innehållet i elevernas utbildning genom att de påverkar elevernas lärande och värdering av olika kunskaper. Amerikansk och brittisk erfarenhet visar att nationella prov dessutom påverkar (i större eller mindre utsträckning beroende på vad resultaten är avsedda att användas till) lärarnas val av stoff och undervisningsformer (Jones et. al., 1999 och Stiggins, 1998).

Metod

Ämnet kunskapsbedömning är stort och spänner över flera discipliner och jag har många gånger känt att jag borde ha haft en djupare kunskap om det "redan från början" för att kunna orientera mig i den enorma litteraturen och finna en bra struktur för mitt eget arbete. Nu har min egen kunskapsutveckling kommit att ske lika mycket under "skrivfasen" som under "läsfasen", vilket innebär att jag under hela arbetet har sett det som nödvändigt att läsa nytt och läsa om, skriva om och skriva nytt. Min förståelse av ämnet, liksom min uppfattning om vad en sådan här text bör handla om och hur detta bäst gestaltas, har alltså förändrats under skrivprocessen. Jag känner fortfarande inte att denna process har "stannat". Men någon gång måste man sätta punkt.

På en konkret nivå har jag gått till väga så att jag har sökt litteratur i databaserna Libris och ERIC (för söktermer som använts vid sökning i LIBRIS, se bilaga 1). För sökning i den pedagogiska databasen ERIC har den engelska söktermen "assessment" använts. Med utgångspunkt från ett urval av dessa skrifter styrt av arbetets syfte och teoretiska inriktning och från den kunskap, som jag erhållit genom tidigare arbeten, om vilka verk som betraktas som klassiska eller representerar vissa inriktningar, sökte jag vidare efter relevanta texter i de valda böckernas referenslistor. Den litteratur, som ligger till grund för arbetet utgörs allmänna, "orienterande" böcker i området, böcker av handbokskaraktär med inriktning mot tillämpning av olika typer av bedömning, forskningsrapporter, sådana debatterande, kritiska artiklar och böcker, som utgör inlägg i den vetenskapliga diskussionen kring bedömning, samt av litteratur från angränsande områden som jag har funnit relevanta för arbetet.

Urvalet av texter har alltså inte följt en strikt systematik och andra skrifter kunde mycket väl ha haft sin plats i framställningen. En strävan har dock varit att urvalet skall spegla de huvudsakliga perspektiv på kunskapsbedömning som återfinns i forskningslitteraturen.

Arbetet med denna skrift har naturligtvis innefattat en rad forskningsmetodiska beslut. Dessa har strukturerats av ställningstaganden på epistemologisk och ontologisk nivå. På den ontologiska nivån har utgångspunkten varit att materiella villkor och maktförhållanden strukturerar sociala praktiker och förståelseformer och vice versa, och att dessa förhållanden aldrig är statiska. Det innebär att de materiella och politiska strukturerna i ett samhälle vid en viss historisk tidpunkt sätter gränser för vad som kan tänkas, uttryckas och accepteras, och gör vissa tankeformer och sociala praktiker mera troliga och förnuftiga än andra i detta samhälle, samtidigt som tanke och handling aldrig är helt determinerade, eftersom människor är kreativa och världen så komplex, att alternativa perspektiv trots allt kan tas. Detta ger möjlighet till genomsådande av illegitim makt och därmed till motstånd, som i sin tur kan leda till förskjutningar i maktrelationerna och förändringar i de materiella villkoren. En utgångspunkt på den epistemologiska nivån är, att människors förståelse av världen i stor utsträckning är medierad genom språket, vilket innebär att de språkliga kategorier som en individ eller ett samhälle förfogar över strukturerar tänkandet. Merleau-Ponty menar att en människas perception av verkligheten alltid är en blandning av det iakttagande subjektet och objektet, och att objektiv kunskap om "världen som sådan" därför i princip är omöjlig. I detta arbete är fokus just på olika förståelseformer, deras rationalitet och maktanspråk i förhållande till samhället och historien. I sin bok "Diskursanalys som teori och metod" (2000), skriver Winther Jørgensen och Phillips, att diskursanalysen inte är ute efter att frilägga en objektiv sanning, utan att dekonstruera och kritiskt granska olika diskurser.²

² Diskurs definieras här i vardagliga termer, som "ett bestämt sätt att tala om eller förstå världen eller ett utsnitt av världen" (Winther Jørgensen & Phillips (2000).

I en mycket allmän mening kan man se stora delar av föreliggande arbete som en form av diskursanalys, då en viktig del i det har varit att ”dekonstruera” vetenskapliga texter om bedömning för att få syn på underliggande ideologiska och teoretiska mönster. ”Diskursanalys” är dock ett vitt begrepp, som omfattar flera olika inriktningar, som på olika punkter skiljer sig åt ifråga om såväl teori som konkret metod. Winther Jørgensen & Phillips (ibid.) framhåller ändå fyra ståndpunkter, som delas av olika diskursanalytiska angreppssätt

- Språket är inte en avspeglning av en redan existerande verklighet.
- Språket är strukturerat i mönster eller diskurser – det är alltså inte tal om ett generellt betydelsesystem som i den saussureska traditionen utan om flera system där betydelserna skiftar från diskurs till diskurs.
- Dessas diskursiva mönster bevaras och förändras i diskursiva praktiker.
- Bevarandet och förändringen av mönstren ska därför sökas i de konkreta kontexter där språket sätts i spel.

I detta arbete har inte de teoretiska grunderna för urval, analys och presentation motiverats i förhållande till någon specifik diskursanalytisk inriktning, men de bygger på de ovanstående gemensamma utgångspunkterna.

Vilka kriterier är då relevanta för att värdera validiteten eller giltigheten i ett arbete som detta? Det kommer naturligtvis an på vilka anspråk man gör på generaliserbarhet och hur man ser på forskningsobjektet. I detta arbete, där forskningsobjektet utgörs av diskurser, som alltså inte ses som generella betydelsesystem, blir validiteten inte så mycket en fråga om representativiteten i och omfattningen av urvalet av texter, som om att

- en mångfald av perspektiv på de fenomen som fokuseras finns med
- att tolkningsramarna för de texter som används beskrivs och relateras till de sammanhang där de är i bruk

- att begrepp som framstår som kritiska (betydelsefulla) analyseras som en del i de språkliga praktiker där de ingår.

Framställningen ska med andra ord innehålla så många och så tydliga referenser till ursprungstexterna och vara så logiskt konstruerad att det är möjligt för läsaren att pröva alternativa tolkningar och dra andra slutsatser. Det har varit min strävan att, utan att veckla in mig i så långa och många begreppsanalyser, att texten blir otymplig och detaljerna skymmer helheten, uppnå en sådan tydlighet. I vilken mån jag har lyckats att balansera kraven på tydlighet och noggrannhet med kraven på helhet och läsbarhet är till sist läsarens bedömning.

Kunskapsbedömning i ett historiskt perspektiv

- När började man systematiskt att bedöma elevers kunskaper?
- Mot vilka samhälleliga behov svarade institutionaliseringen av kunskapsbedömning i skolan?
- Vad har utbildningens förändrade samhällsvillkor inneburit för sättet att se på, organisera och genomföra bedömning?

Den institutionaliserade bedömningen av elevers kunskaper har fyllt olika funktioner i samhället vid olika tidpunkter. Omvänt kan man se att de sätt på vilka man vid en given tidpunkt ser på, organiserar och genomför bedömning inom utbildningssystemet har att göra med de samhälleliga behov som bedömningen och utbildningen för tillfället har att svara mot. Att lärare och beslutsfattare har insikt om de historiska och sociala betingelserna för bedömningsfilosofi och -praktik under olika perioder är viktigt av den anledningen att den bedömning som görs av elever i skolan har betydelse inte bara för deras framtida karriärmöjligheter, utan sannolikt också för deras syn på kunskap, samhället och sig själva som medborgare och individer. I denna första del ska jag redogöra för vilka motiv för institutionaliserad kunskapsbedömning som har varit viktiga vid olika tidpunkter.

De första institutionella bedömningarna

Som det första nationella examinationssystemet brukar anges det kinesiska. Det var redan på 200-talet som man i Kina började använda sig av formella examinationer i meningen skriftliga, offentliga, konkurrensinriktade prov (se t ex Eckstein & Noah, 1993; Gipps & Murphy, 1994 och Madaus & O'Dwyer, 1999). Proven användes vid urval av sökande till högre anställningar inom statlig och militär förvaltning. Sådana anställningar var förknippade med höga sociala positioner och hade tidigare varit reserverade för mandarinernas, de statliga ämbetsmännens, söner.

När de offentliga examinationerna infördes var det med syftet att dessa privilegierade positioner i princip skulle göras tillgängliga för unga män ur alla samhällsklasser. I praktiken blev det dock ingen större förändring, utan de åtråvärda tjänsterna kom även fortsättningsvis nästan undantagslöst att besättas av mandarin-söner. Detta berodde på att de kunskaper som testades i proven nästan uteslutande var av ett slag som var förknippade med mandarinernas erfarenheter och intressen – i praktiken var det därför nästan omöjligt för unga män med annan bakgrund att hävda sig i konkurrensen med mandarinerna (Madaus & O'Dwyer, 1999). De sökandes kunskaper testades inom flera områden, bl a militär skicklighet, och det mest prestigefyllda provet inbegrep bl a argumentationsanalys, retorik och poetik. Denna del kom dock så småningom att utgå, eftersom det ansågs alltför svårt att uppnå en tillfredsställande objektivitet vid bedömningen (ibid.). I princip överlevde dock det kinesiska examinationssystemet ända in på 1800-talet (Eckstein & Noah, 1993).

I Europa har institutionaliserade former av kunskapsbedömning existerat sedan 1100-talet, då universiteten i Paris och Bologna var de lärosäten som introducerade examinationer i vår mening. Dessa bestod av muntliga förhör, där studenterna skulle återge inlärdas svar på kända frågor, som ofta bestod av dilemman (Madaus & O'Dwyer, 1999). En annan form av institutionalise-

rad bedömning vid den här tiden var hantverksskrånas gesällprov.

De första egentliga examinationssystemen i Västeuropa uppstod dock, enligt Eckstein och Noah (1993) inte förrän slutet av 1700-talet. Då de moderna staterna industrialiserades och byggde ut sina kommunikationer och sin byråkrati, blev det allt vanligare att använda sig av officiella, skriftliga prov för selektion av anställda. Det som skilde dessa bedömningssystem från tidigare var att de byggde på en kvantifiering av kunskaper. Innan industrialismen hade all kunskapsbedömning skett uteslutande i kvalitativa termer (Madaus & O'Dwyer, 1999). Vid 1800-talets mitt hade de kvantitativa proven fått stor spridning. Det kinesiska systemet anses, trots den tidsmässiga överlappningen, inte ha haft något nämnvärt inflytande på dessa europeiska examinationssystem (Eckstein & Noah, 1993).

Vad var då syftet med de institutionella bedömningarna? Medan det kinesiska systemet hade varit helt inriktat på *selektion* (urval), var de europeiska lika ofta inriktade på *kvalifikation*, d v s att utfärda legitimation för vissa yrkesgrupper och att ge tillträde till högre utbildning. Yrkesutbildningen, som förr huvudsakligen hade skett på arbetsplatsen, genom deltagande och successivt ökande ansvar, ersattes efterhand med formell utbildning, som resulterade i mera "portabla" och flexibla kvalifikationer i form av yrkesexamen eller betyg (Broadfoot, 1996).

Hur motiverades införandet av offentliga examinationer? Uppkomsten av dessa offentliga, storskaliga och *kvalificerande* examinationssystem förknippas med 1800-talets kombination av demokratiska samhällsrörelser och begynnande kapitalism (Broadfoot, 1995). De motiverades dels utifrån de kompetenskrav som nya och alltmer specialiserade arbetsuppgifter inom den växande industrin och byråkratin ställde, dels ideologiskt – de var ett sätt att garantera "rättvisa" och minska de vanligt förekommande anklagelserna om nepotism, mutor och favorisering av elever från välbärgade hem (ibid.).

Hur påverkades utbildningens samhällsstatus genom införandet av offentliga examinationer? Examinationssystemens genomslag innebar en institutionalisering av förhållandet mellan utbildning och social position, och en klass av intellektuella började växa fram. Den bildningskultur som dessa kom att dela, kom, menar Broadfoot (1996), efterhand att bli de intellektuellas gemensamma signum. De kulturella tillgångar, som alltså förvärvats genom utbildning och på vilka examina utgjorde bevis, öppnade i sin tur dörrarna för yrkeskarriärer som ledde till höga sociala positioner. Att de intellektuella började framträda som en egen klass och distingera sig från övriga grupper genom sina kulturella normer och vanor, medförde emellertid en försvagning av examinationssystemets potential att likställa individer med olika social bakgrund i konkurrensen om samhällets resurser. I och med att examinationssystemet både definierade och ”förräntade” den nya klassens ”kulturella kapital”³, så blev det också till ett led i den sociala reproduktionen. Jämförelsen med mandariner-na ligger nära till hands, och har ibland också använts som en metafor för att beskriva den bildade klassens exklusiva förhållande till utbildningssystemet (Broadfoot, 1996; Torrance, 1995). Det feodala systemet där sociala privilegier ärvdes inom familjen, började alltså under denna tid ersättas av ett system med utbildningssystemet som en förmedlande länk och den sociala reproduktionen hade kopplats till den kulturella (Åberg, 1992).

I majoriteten av de västeuropeiska länderna överfördes examinationssystemet för den högre utbildningen så småningom till ungdomsskolan, vilket innebar att elevernas avgångsbetyg, istället för att som tidigare grundas enbart på lärarnas bedömning, i hög grad kom att baseras på resultaten av offentliga examinationer i de olika ämnena (Calfée & Masuda, 1997). Detta stärkte betygens allmänna legitimitet:

3 Begreppet ”kulturellt kapital” myntades-, av den franske sociologen Pierre Bourdieu och beskrivs av Jönsson m fl (1994) som: ”en slags kulturell kompetens som utgörs av tillgångar i form av språkbruk, kunskaper, erfarenheter, värderingar, sätt att uppfatta och tänka, tala och röra sig mm och utgör en art av symboliskt kapital”.

Arbetsgivare, utbildningsadministratörer, skattebetalare, föräldrar, politiker – alla kände sig mera tillfreds när de visste att avgångsbetygen baserade sig på resultaten av formella prov i de olika ämnena.
(Eckstein & Noah, 1993, s. 4).

Vad betydde då dessa examinationer för undervisningen? I den utsträckning som examinationerna institutionaliserades kom de naturligtvis också att påverka innehållet i skolan. Examinationerna kan rentav ses som själva upphovet till läroplanerna, menar Eckstein & Noah .

Förekomsten av examinationer innebar att en strukturerad kunskapsmängd måste definieras, skrivas ner och publiceras, ämne för ämne, så att den kunde läras ut i skolorna. Tydligare mål och en mera strukturerad planering av undervisningen ökade utbildningen synbara värde. Samtidigt tenderade dock det urval av kunskaper som testades i examinationerna att bli definierande för vad som lärdes ut i skolorna, vilket ledde till det vanliga klagomålet att undervisningen domineras av proven.
(ibid., s. 17)

Hur såg det ut i Sverige? I det allmänna skolsystemet var det bara realskola, läroverk och flickskola som avslutades med formella, centralt utformade examinationer. I folkskolan förblev bedömningen och betygssättningen av eleverna de enskilda lärarnas ansvar. Visserligen reste på ecklesiastikministeriets uppdrag skolinspektörer runt i landet och kontrollerade elevernas kunskaper i de föreskrivna momenten genom muntliga förhör. Syftet med denna kontroll var emellertid att utvärdera undervisningens kvalitet vid de olika skolorna och inte att sätta betyg på de enskilda eleverna (se Lundgren, 1979).

I Wales och England infördes en liknande statlig inspektion av "elementary schools" i slutet av 1800-talet, med anledning av att utbildningskvaliteten uppfattades som så låg att staten såg sig föranledd att allokera resurser för att effektivisera undervisningen (Eckstein & Noah, 1993). Precis som i den svenska folkskolan gick inspektionen till så, att eleverna förhördes muntligen av skolinspektörerna. Detta förhör kom så småningom att formaliseras till en typ av nationellt prov. Såväl i England och Wales som i Sverige undgick privatskolor och hemundervisning denna typ av statlig kontroll. Inom dessa skolformer ansågs

länge de lokala prov och betyg som lärarna själva gav, ge tillförlitliga mått på elevernas kunskaper. I de brittiska länderna infördes dock år 1917 den för elever ur alla skolformer gemensamma examination som senare utvecklades till den nuvarande GCSE⁴ (ibid.).

Både i England och USA användes (och används på vissa håll fortfarande) elevernas betyg som en indikation på lärarnas effektivitet och skolornas kvalitet. Som sådana kunde de utgöra grund både för enskilda lärares lönesättning och för ekonomisk tilldelning till skolorna. Dessutom publicerades betygsstatistik från olika skolor med avsikten att vara de föräldrar behjälpliga som ville välja skola för sina barn (Broadfoot, 1996; Elliott, 1998).

⁴ GCSE är en förkortning av General Certificate of Secondary Education och betecknar den examen som alla elever i Storbritannien genomgår vid c:a 16 års ålder som avslutning på sina studier i *secondary school*. GCSE kallas ibland också 16+.

Intelligenstestningens tidsålder

Med 1900-talet inleddes en period som brukar benämnas "the testing age" (Broadfoot, 1995; Eckstein & Noah, 1993; Rosén, 1998). Vid denna tidpunkt fanns en utbredd tro i samhället på uppfattningen att varje individ föds med ett viss mått av inneboende intelligens. Intelligensten sågs om en tämligen stabil egenkap hos individen och i princip opåverkbar genom t ex utbildning. Man var därtill övertygad om att intelligensen gick att mäta, och flera framträdande psykologer utvecklade instrument med vars hjälp man med vetenskaplig precision och objektivitet skulle kunna fastställa individers intelligens.

Hur ska man då förstå det stora intresse som växte fram vid sekelskiftet för att mäta människors intelligens? En del av bakgrunden var att resultaten från intelligensmätningar antogs utgöra tillförlitliga mått på olika individers duglighet för yrke och studier, och därmed kunde användas för att predicera individers framtid (Broadfoot, 1996; Eckstein & Noah, 1992; Gipp & Murphy, 1994). Dessa tankar stämde väl överens med den individualistiska syn på utbildning, som var dominerande i västvärlden vid denna tidpunkt, menar Broadfoot (1996). Det faktum att Alfred Binet i Frankrike, Charles Spearman i England och Lewis Therman i USA oberoende av varandra och i stort sett samtidigt (alldeles i början av seklet) presenterade var sitt intelligenstest, kan ses som ett tecken på att den ökade sociala skiktning som skett till följd av den tilltagande differentieringen i arbetslivet, givit upphov till ett samhälleligt behov av vetenskapligt accepterade metoder för att sortera människor för olika sociala positioner (även om intelligenstesterna inte alltid tagits fram i detta syfte).

Rosén (1998) sammankopplar intelligensforskningens stora genomslag i början av 1900-talet med ett behov av förmodat mera "objektiva" metoder för att diagnostisera och rangordna människor, något som dittills alltså hade skett på vad som framstod som mera subjektiva grunder. "Objektivitet" kan i detta sammanhang ses som synonymt med den precision och

värderingsfrihet, som förknippas med vetenskapliga forskningsideal. Att rangordna människor på basis av deras resultat på olika typer av intelligenstester tycks alltså inte ha betraktats som särskilt kontroversiellt i sig ur moraliskt eller ideologiskt hänseende, utan verkar främst ha setts som en metodteknisk fråga.

Eckstein & Noah (1993) ser testningens genombrott tillsammans med idén om objektiva, konkurrensinriktade examinationer som en del av den filosofi, som såg det sociala arvet som i princip underkastat samma lagar som det biologiska, vilket innebär att samhället som helhet skulle gynnas av att dess resurser framför allt kom de dugligaste individerna del; i filosofin ingick också den moraliska kodexen att de dugligaste skulle få vederbörligt erkännande för sin överlägsna förmåga och arbetsinsats. Examinationerna och testerna utgjorde då, ansågs det, instrument med vars hjälp människors varierande duglighet för samhället skulle kunna fastställas:

Den utbredda användningen och livliga tillväxten av examinationssystem kan ses som ett naturligt komplement till en tro på individens autonomi och ansvar och en tro på att ett rättvist samhälle belönar var och en på basis av hans eller hennes fastställda mentala egenskaper. (Eckstein & Noah, 1993, s.17)

Intelligensideologiernas genomslag hade troligen också en religiös bakgrund:

the belief that some are predestined for salvation and others for damnation is not only compatible with the belief that only some are chosen intellectually, but more significantly that such gifts might properly be used as an arbiter of individual merit and worth. (Lohman, 1997, s. 68).

Intresset för intelligensforskning och studier av individuella differenser har varit betydligt större i länder som dominerats av puritanism och kalvinism, menar Lohman (ibid.) än vad som varit fallet i andra länder. Lohman menar att dessa religiösa och ideologiska strömningar utgör den kanske viktigaste bakgrunden till det tidiga 1900-talets stora och snabba utbredning av konkurrensinriktade och "objektiva" prov i utbildningssammanhang. Men också den protestantiska rörelsen där den starka ton-

vikten på individens karaktär och handlingar var grundläggande (Dahl, 1999), bör ha utgjort en passande värdegrund för intelligensideologierna.

Den omfattande testningen under seklets första hälft motiverades emellertid också utifrån de politiska liberala och socialdemokratiska idéer om jämlikhet, som började göra sig gällande i och med industrialiseringen. Också där finns en fokusering på individen. Enligt dessa idéer ska likhet i möjligheter råda mellan individer att realisera sina livsplaner på så sätt, att en persons livschanser är en funktion av dennes förvärvade egenskaper (t ex utbildningsmeriter) och inte de sk tillskrivna (t ex det sociala ursprunget) (se t ex Åberg, 1992).

Ett utbildningssystem som alla medborgare har tillträde till och där de behandlas och bedöms på lika villkor är naturligtvis viktigt för att människor med olika bakgrund ska ha chans att förverkliga sina intressen och göra bruk av sina gåvor, något som för individen både har ett värde i sig och en koppling till karriärmöjligheter och social status. Broadfoot (1996) beskriver de statliga examinationssystemen som viktiga redskap i strävandena mot att förverkliga det ”meritokratiska samhället”:

The certification process is indeed the epitome of the apparently meritocratic basis of contemporary society, since in theory it allows free competition based on academic ability and industry and thus is regarded as the fairest basis of the allocation of opportunities for high status or remunerative careers. (ibid., s. 32).

Även om intelligenstesterna inte utvecklades med syftet att utgöra en del av det reguljära examinationssystemet (Spearman utvecklade sina tester för att mäta intelligensen hos värnpliktiga med syftet att sätta ”rätt man på rätt plats” och Binet sitt för att kunna diagnostisera elever med inlärningssvårigheter), så kom de att få ett avgörande inflytande på den allmänna uppfattningen av intelligensens betydelse. Catherine Taylor (1994) menar att det sätt att se på och genomföra kunskapsbedömning som under lång tid har dominerat kunskapsbedömningen väster-

ländska samhällen, nämligen ”mätmodellen”⁵ baserar sig på samma föreställningar om intelligentsetningen, nämligen

1. Människor skiljer sig från varandra i fråga om olika slags begåvning.
2. En individs mängd av ett visst slag av begåvning kan beskrivas i relation till fördelningen av andra liknande individers mängd av detta slags begåvning.
3. Man kan utveckla instrument som tillförlitligt mäter sådana individuella differenser i fråga om begåvning.

(Taylor, 1994, s. 237)

Då man i utbildningssammanhang har använt metoder som bygger på teorin om begåvning och individuella differenser för att fastställa tillförlitligheten hos prov, tolka provresultat och för att skapa förtroende för testresultat som en indikation på individers ”sanna” förmåga (ibid.), har dess underliggande antaganden om begåvning byggts in utbildningssystemet – trots att kunskapsbedömning i utbildning ju handlar om att få syn på lärandets processer och resultat, snarare än att söka efter mått av statistiska egenskaper.

”Mätmodellens” genomslag hade också en praktisk bakgrund: på 50-talet uppfanns en optisk scanner som gjorde det möjligt att snabbt rätta stora mängder av prov, som bestod av en uppsättning frågor med fasta svarsalternativ, d v s så kallade *multiple-choice*, eller flervalssprov (de kallades även ”objektiva prov”) (Hambleton & Sireci, 1997). Formen med standardiserade flervalss frågor som kunde rättas mekaniskt kom snart att tas i vitt bruk inom det amerikanska utbildningssystemet. Den nya tekniken gjorde det möjligt att distribuera samma prov och bedöma elevernas kunskaper med samma kriterier över hela landet.

5 Taylor lanserar begreppet ”mätmodellen” för att beskriva just den typ av prov som bygger på att uppgifterna är av ett sådant slag och tillräckligt många för att de ska åstadkomma en spridning av eleverna, ha ett för eleverna okänt innehåll, på att provsituationen är standardiserad och att bedömningskriterierna hålls hemliga för eleverna.

Därmed fick staten ett medel för att kontrollera vad som lärdes ut på lokal nivå och makten över innehållet i den konkreta utbildningen kom på sätt att centraliseras (Hambleton & Sireci, 1997). I USA hade fram till mitten av 1950-talet bedömningen av eleverna nästan uteslutande varit de undervisande lärarnas ansvar – i och med denna förskjutning minskade lärarnas makt över innehållet drastiskt med följderna att kunskapsbedömningen även på klassrumsnivå kom att bli starkt präglad av de nya flervalssprovens innehåll och form (ibid.). Enligt Madaus & O'Dwyer (1999) har dock flervalssproven och den typ av storskaliga utvärderingar som syftar till central kontroll av innehållet, aldrig haft tillnärmelsevis så stort genomslag i Europa som i USA. I europeiska länder, menar dessa författare, sker huvuddelen av bedömningen av eleverna på basis av prov som utformats av de undervisande lärarna och essäprov är vanligare än flervalssprov.

I varierande grad har olika stater betraktat det som problematiskt, att likhet i formella rättigheter inte upphäver skillnaderna mellan individer med olika social bakgrund vad gäller t ex övergångsfrekvensen till högre studier, och i varierande grad har man försökt minska skillnaderna genom olika kompensatoriska åtgärder. I Sverige har stora forskningsresurser ägnats just åt att reda ut orsakerna till den sociala snedrekryteringen till högre studier och att föreslå och utvärdera olika statliga insatser för att minska snedrekryteringen (Husén, 1999). Kjell Härnqvist fastslog 1957 att det fanns en oproportionerligt stor andel studiebegåvade ungdomar från arbetarklassen i Sverige, som avstod från högre utbildning. Denna grupp benämndes "begåvningsreserven" (ibid.). Den stora begåvningsreserven inom arbetarklassen var problematisk både ur ideologiskt och ekonomiskt hänseende; den gjorde det svårt att hävda att alla medborgare inte bara i teorin, utan också i verkligheten, faktiskt konkurrerade på lika villkor om sociala positioner, och den visade att en god del av den intellektuella kapaciteten i landet inte utnyttjades. I Sverige har alltså intelligenstestningen från 1950-talet och framåt använts för att analysera bakgrunden till och föreslå åtgärder för att motverka den sociala snedrekryteringen till högre studier. Senare har intelligensforskning även använts för att argumentera mot essentialistiska och ensidigt biologiska för-

klaringar till könsbundna beteenden och samhällsliga maktstrukturer (t ex Rosén, 1998).

Men intelligensforskning har också använts för att legitimera konservativa ideologier; ett ökänt exempel utgör den moderna bästsäljaren "the Bell Curve" (Herrnstein & Murray, 1994), där resultat från intelligensforskning används för att visa hur den sociala stratifieringen i USA motsvaras av intelligensskillnader mellan olika grupper⁶. Liknande slutsatser har under intelligensforskningens historia dragits av en mängd forskare, där kvinnor, socialt marginaliserade och vissa ras- och etniska grupper har utpekats som intellektuellt underlägsna.

⁶ Författarna har blivit hårt kritiserade av forskare runt om i världen för att ha byggt sina resonemang på invalida forskningsresultat och för att deras slutsatser ofta saknar empiriskt stöd. Att "the Bell Curve" ändå har sålts i rekordstora upplagor och debatterats i alla stora dagstidningar världen över, indikerar att de ståndpunkter som författarna framför inte enbart har avfärdats som dålig forskning eller propaganda, utan har funnits värda att diskutera, både inom forskarsamhället och av "en läsande allmänhet", vilket ju tyder på att övermänniskoidealet inte var helt förkastat i 1990-talets samhällssyn.

Efterkrigstidens utbildningsexpansion

Återhämtningen efter andra världskriget innebar i alla västländer en snabb utveckling av industrin och stor ekonomisk tillväxt. De förbättrade statsfinanserna användes dels för att bygga ut utbildningssystemet, så att industrins växande behov av kvalificerad arbetskraft kunde mötas, dels för att finansiera allmänna system för t ex vård, barn- och äldreomsorg, pensioner, social- och sjukförsäkringar (se t ex Ball & Larsson, 1989). Stora summor satsades på att öka volymerna i den högre utbildningen, som tidigare endast hade varit aktuell för en liten bråkdel av befolkningen⁷. Expansionen var både ideologiskt och ekonomiskt betingad. Å ena sidan sågs ett utbildningssystem som omfattade alla och gav alla lika chans att konkurrera om meriter som grundläggande för demokratin. Å andra sidan innebar, som redan nämnts, den starka tillväxten inom industrin en ökad efterfrågan på utbildad arbetskraft.

Vilken roll spelade då examinationssystemet i uppbyggandet av de moderna välfärdsstaterna? För att fördela platserna i den snabbt växande högre utbildningen på ett rättvis och objektivt sätt, behövdes en omfattande och systematisk testapparat. Det var i samband med detta som det verkliga genombrottet för nationella examinationssystem med direkt koppling till utbildningssystemet skedde.

Denna tidsperiod, som inleddes med sekelskiftet och som nu ibland antas vara passé eller på väg att upplösas, förknippas ofta med begreppet modernitet; man talar om ”det moderna projektet” och ”det moderna samhället”⁸. Det moderna samhället

⁷ Av dem som föddes i Sverige 1905 övergick ca 2% till motsvarande dagens högskoleutbildning, av födda 1950 ca 20% (Erikson & Jonsson, 1993).

⁸ Madaus och O'Dwyer (1999) placerar in ”moderniteten” mellan år 1900 och 1980, och benämner den period som inleds med 1980-talet postmodernitet. Också Haag (2001), svensk etnolog, menar (med hänvisning till Fornäs) att brytningen mellan det ”gamla” förmoderna och det moderna sker omkring 1900, och sätter 1950-talet som starten för en period som benämns ”senmodernitet”. Det är dock inte fråga om några entydiga gränser, och det förekommer att man hänvisar till franska revolutionen som en startpunkt för det ”moderna” samhället.

präglades av en stark framtidstro och utvecklingsoptimism och synen på utbildning och kunskap var pragmatisk, menar t ex Popkewitz (1990). Englund sammanfattar det svenska skolsystemet under denna period:

Från att under 1800-talet ha präglats av halvtidsläsning, ambulerande skolor etc. tog folkskolan relativt fast form under gemensam liberal och socialdemokratisk ledning i och med skolreformen 1919, då folkskolan fick en ny, framåtsyftande underdervisningsplan. Folkskolan behöll emellertid länge sin patriarkaliska – religiöst underdånighetsfostrande och fosterländska inriktning, för att först under efterkrigstiden byggas på vetenskapligt-rationell grund och upprätta en ny balans. Det gällde nu å ena sidan att förbereda sig för en ekonomi i tillväxt, d v s arbetskrafts-förberedelse, och å andra sidan för deltagande i den svenska demokratin. Demokratispekten och synen på skolan som jämlikhetssyftande dominerade under de första efterkrigsåren för att avlösas av ekonomiaspekten under 1950- och 60-tales rekordår: skolan i det föränderliga samhället. (Englund, 1996, s. 27)

Frykman (1999) menar att skolan under denna tid för den enskilde mer än något annat kunde representera en potentiell social språngbräda. För samhället var det alltså en institution som dels producerade den ”know-how”, som behövdes för att säkra den ekonomiska tillväxten och utveckla välfärdssamhället, dels skulle fostra de uppväxande generationerna till demokratiska medborgare.

Objektivt, rättvist, tillförlitligt

Eftersom betyg har den mycket viktiga rollen att kunna ”öppna och stänga dörrarna för individers framtida livschanser” (Broadfoot, 1996, s. 32), har examinationernas viktigaste roll i det moderna samhället varit att reglera konkurrensen om sociala positioner. Att proven kunde identifiera ”de duktigaste” var därför på ett självklart sätt överordnat andra tänkbara syften med bedömningen av elevers kunskaper – exempelvis att identifiera förkunskaper eller diagnostisera inlärningssvårigheter.

Vilka konsekvenser fick betoningen på detta syfte för provens utformning? En rad olika karakteristika gjorde alltså formen för

de tester som utvecklats för att mäta intelligens särskilt lämplig för att bedömningen av elevers kunskaper. Central utformning och distribution möjliggjorde jämförbara resultat för elever från olika skolor och delar av landet, multiple-choice-formatet, som möjliggjorde mekanisk rättning hade ekonomiska fördelar och var objektivt, såtillvida att det garanterade lika behandling av alla som gjorde provet. Dessa kvaliteter bidrog till att prov av denna typ kunde anses uppfylla de höga krav, som samhället med tanke på den kritiska roll de spelade för individernas framtida livschanser, ställde på examinationerna att vara rättvisa och objektiva.

Tiden efter andra världskriget övergick många europeiska länder från starkt differentierade skolsystem (där utbildningen för det stora flertalet bestod av ca sex år i folkskola, de besuttnas barn gick i privata skolformer och endast en liten elit gick vidare till universitet), till någon form av enhetsskola där den huvudsakliga differentieringen inte skedde förrän inför gymnasiestudierna (Broadfoot, 1996; Åberg, 1992). Detta resulterade i att den allmänna utbildningsnivån i dessa länder höjdes – fler och fler gick i skolan allt längre och fler och fler fick därigenom också de kvalifikationer som tidigare varit förbehållna fåtallet. Höjningen av den allmänna utbildningsnivån ledde till att de genom utbildning förvärvade kvalifikationerna delvis förlorade sin betydelse som urvalsinstrument (Åberg, *ibid.*). Därför måste examinationernas kvalificerande funktion i allt högre grad kombineras med en sorterande funktion.

1970-talets utbildningsreformer

I början av 1970-talet blev utbildningssynen i västvärlden mer radikal, något som naturligtvis kan sättas i samband med den allmänna tidsanda som studentrevolten 1968 både kan ses som uttryck för och del i uppkomsten av. SIA-reformen som genomfördes i den svenska grundskolan var ett uttryck för socialdemokratiska strävanden att skolan skulle utjämna skillnader i livschanser och ge en medborgarfostran byggd på en aktiv demokratisyn, Richardsson (1999). Med SIA-reformen skulle undervisningen knytas starkare till samhället och föreningslivet och i högre utsträckning göra elevernas vardagserfarenheter till utgångspunkt för skolarbetet. Gränsen mellan manuellt och intellektuellt arbete skulle mjukas upp och undervisningen bli mindre akademisk (Englund, 1996; Larsson & Ball, 1989). I SIA-propositionen betonades också att grundskolan skulle vara sammanhållen och att elever med olika social bakgrund skulle mötas i samma klassrum (Schüllerqvist, 1996). Englund skriver vidare om skolsynen i Sverige på 1970-talet:

Demokratiaspekten och synen på skolan som ett jämlikhetssyftande, fördelningspolitiskt instrument återkom med kraft under 1970-talet (...) Sammantaget kan också påstås att den tidigare dominerande vetenskapligt-rationella inriktningen i vissa stycken överskreds genom en explicit fostran till demokrati och ett konfliktperspektiv. (Englund, 1996, s. 27)

Den ekonomiska nedgången och den politiska högervändning, som sedan ägde rum i olika grad i USA och i de västeuropeiska länderna i början av 1980-talet, tillsammans med en dramatisk nedgång i ekonomin i alla dessa länder, gjorde dock att dessa progressivistiska utbildningsansatser kom av sig – i många länder beskylldes förändringarna i skolan för att vara, om inte själva upphovet till, så åtminstone en väsentlig del av orsakerna till den ekonomiska krisen (Larsson & Ball, 1989). I USA och Storbritannien ledde detta till att den konservativa rörelsen ”back-to-basics” startade, och i Sverige ändrade den offentliga diskussionen om skolan förtecken: man blev mer intresserad av hur effektiv skolan var på att lära ut kunskaper, än av skolan som bildningsinstitution och agent för att öka den sociala rörligheten (ibid.).

Formerna och utgångspunkterna för bedömningen av elevernas kunskaper tycks dock i huvuddelen av svenska klassrum inte ha genomgått några genomgripande förändringar under denna period under 1970-talet då man experimenterade med mera vardags- och samhällsanknuten undervisning⁹. Möjligen förstärktes snarare provens reproduktiva karaktär under 1970-talet genom det genomslag som den så kallade undervisningsteknologin fick i skolorna. Englund (1996) beskriver hur undervisningsteknologi och progressivism under denna period omväxlande och parallellt framhålls som modell för undervisningen. Genom undervisningsteknologin med sina programmerade läromedel kom proven, utöver sin sorterande funktion, att få rollen som pedagogiskt hjälpmedel. Läroböckernas diagnosprov med tillhörande "facit" gjorde att eleverna själva kunde avgöra när de skulle gå vidare till nästa avsnitt och vad de behövde öva mera på. Denna form av individualisering lösgjorde lärartid och ansågs, i enlighet med den behaviouristiska inlärningsteori som var dess bas, öka elevernas motivation. På så sätt ökade läroboksförlagens inflytande över både innehåll och bedömning i skolan på bekostnad av lärarnas.

-
- 9 Här stödjer jag mig på Gunilla Svingbys uppfattning, jag har inte funnit någon empirisk undersökning av detta förhållande. Shepard (2000) beskriver dock, troligen framför allt grundat på den amerikanska erfarenheten, utvecklingen av relationen mellan bedömning och undervisning på följande sätt:
- 1) C:a 1900-2000+: Undervisningen vilade på behaviouristisk inlärningsteori och ärftlighetsbaserade intelligensteorier tillsammans med läroplaner styrda av idéer om ekonomisk effektivitet och kunskapsbedömningen byggde på psykometrisk mätteori
 - 2) 1980-2000+: kunskapsbedömningen förlitar sig fortfarande till övervägande del på mätteori, medan undervisningen alltmer influeras av sociokulturella och kognstuktivistiska teorier
 - 3) 1990.2000+: Ett nytt "paradigm" är under framväxt där nya synsätt på utbildning och kognitiva och konstruktivistiska inlärningsteorier börjar integreras med klassrumsbaserade bedömningsmodeller.

Bedömning i det postmoderna samhället

Också på 1990- och 2000-talet har regeringarna i Europa och övriga västvärlden gjort stora ekonomiska satsningar på utbyggnad av sina utbildningssystem. Dessa satsningar har varit av en aldrig tidigare skådad omfattning – aldrig förr har så stor andel av statens ekonomi använts för utbildning och aldrig har en så stor andel av befolkningen varit i utbildning eller haft högskoleexamen. Exempelvis hade 2001 29 % av Sveriges befolkning någon form av eftergymnasial utbildning (SCB, 2001) och under läsåren 1993/94-2000/01 ökade antalet studenter i högskoleutbildning från c:a 256 300 till 330 200, (SCB, 2002). Denna stora expansion i västvärlden av den högre utbildningen har, menar t ex Broadfoot (1996) dock inte motiverats av arbetsmarknadens behov av utbildad arbetskraft så mycket som av statens behov av att hålla unga människor sysselsatta, då arbetslösheten, särskilt bland unga och särskilt inom okvalificerade yrken, har ökat kraftigt sedan slutet av 1980-talet.

Om högre studier för ett par decennier sedan i det närmaste garanterade kvalificerad anställning, så har de senaste decenniernas kraftiga ökning av platser i högre utbildning ”devalverat” högskoleexamen till ett nödvändigt, men inte tillräckligt urvalskriterium för attraktiva karriärer, menar Åberg (1992). Ett antagande som måste ses som rimligt är att socialt och kulturellt kapital som en följd av denna utveckling har ökat i betydelse som konkurrensmedel, medan betydelsen av akademiska meriter har sjunkit (jfr Bourdieu, t ex 1982). Under den tid då de kvalifikationer som man skaffat sig genom utbildning i det närmaste garanterade en hög status, hade de kunskaper som man förvärvade under utbildningen ett stort och obestridligt bytesvärde och proven kunde därför i hög grad motivera eleverna att studera. Hargreaves (1989), som också pekar på att det har skett en försvagning av sambandet mellan utbildningskvalifikationer och framtida livschanser, menar att denna försvagning har bidragit till en legitimitetskris för skolan, som när kunskapernas byte-

svärde sjunker får svårt att motivera eleverna. Det får som följd att proven allt mindre fungerar som ”morot” för olika kategorier av elever.

Motivera istället för att sortera

Förändringar i den ekonomiska strukturen har alltså påverkat utbildningssystemets roll i förhållande både till samhället och till de individer som utbildar sig. Vilka konsekvenser kan denna rollförändring i sin tur antas ha haft när det gäller bedömningsfunktioner i samhället, klassrummet och för den enskilde? Och om bedömningsfunktioner ändras – vad innebär det för hur man rent konkret går tillväga vid bedömningen?

Frågor om kunskapsbedömning ägnades alltså internationellt stor uppmärksamhet under 1980- och 90-talet. Behovet av att finna nya vägar för att motivera elever – särskilt svagpresterande elever, som man annars skulle riskera att förlora till arbetslöshet och social marginalisering, kan vara en del av förklaringen till att lärare, politiker och utbildningsforskare på senare tid har sökt så intensivt efter nya former och funktioner för kunskapsbedömning. Mot denna bakgrund är det logiskt, att dagens diskussioner om prov inte handlar så mycket om sortering och konkurrens, som om ”djupinläring”, självförståelse och individualiserad undervisning:

... just as psychometrics and norm-referenced assessment generally were the reflection of this emphasis on selection, so competency-based and criteriareferenced assessments are a reflection of the economic characteristics of contemporary ”post-industrial ” societies in which there is a pressing need to encourage more and more, rather than reduce to fewer and fewer, learners at every level of the educational system.

(Broadfoot, 1996: 49)

När reaktionen mot den i positivism och behaviourism grundade bedömningspraxis, som ju spelat en så central roll i de moderna massutbildningsinstitutionernas utveckling började komma i slutet av 1980-talet i de flesta västländer, så kan det förstås mot bakgrund av dels de ovan skisserade förändringarna

i den ekonomiska strukturen (övergången till det stadium av kapitalismen som ibland kallas senkapitalism, t ex Broadfoot, 1996), dels förändringar gällande bl a kunskaps- och utbildningssyn, som på ett eller annat sätt kan förknippas med vad som mycket övergripande brukar beskrivas som övergången från det moderna till det postmoderna samhället¹⁰.

Ny kunskapssyn

En genomgripande sådan förändring gäller kunskapssynen inom samhälls- och humanvetenskaperna. Sedan slutet av 1960-talet har den logiska empirismen och positivismen blivit starkt ifrågasatt som grund för kunskap inom dessa discipliner (Alvesson & Sköldberg, 1994). Den inomvetenskapliga debatten har sedan dess intensifierats, och de senaste tio åren kan ståpunkten, att vetenskap som rör mänskligt handlande behöver metoder och teorier, som undersöker människors tänkande och rationalitet inte kan utgå från samma tekniska kunskapsintresse som naturvetenskaperna, knappast uppfattats som kontroversiell. När det gäller pedagogisk forskning, skriver t ex Eisner (1991) att det rentav har skett ett paradigmskifte. Detta innebär i så fall inte att t ex den forskning inom pedagogiken, som bygger på intelligenstestning, har förkastats eller upphört men väl att de kvalitativa forskningsansatser som är inspirerade av antropologiska arbetssätt och utgår från t ex hermenentiska och konstruktivistiska kunskapsbegrepp har utvecklat sina metoder och sin filosofiska grund och att de har en vetenskaplig legitimitet som de inte hade förut. Detta har bl a lett till att innebörden i centrala vetenskapliga begrepp som validitet har kommit att utmanas

¹⁰ Ordet "postmodern" kan referera till en tidsepok (som ibland anses inledas efter andra världskriget, ibland så sent som på 1980-talet), en intellektuell, estetisk eller litterär stil förknippad med denna tidsepok, eller en epistemologisk hållning – när det gäller det sista, så används också begreppet "postrukturalism" med samma eller liknande betydelse som "postmodernism". Begreppet "det postmoderna samhället" används ofta synonymt med "det senkapitalistiska samhället" och ses då alltså som ett led i en samhällsekonomisk process (Alvesson & Sköldberg, 1994). Jämför t ex Hargreaves (1999) i boken "Läraren i det postmoderna samhället".

och ges nya tolkningar (se t ex Kvale, 1989 och Messick, 1993). Överhuvudtaget har den kvalitativa forskningens ökade utbredning och självkänsla lett till metodologisk och epistemologisk reflexion inom olika discipliner.

Intresset har således ökat markant för metoder, som fokuserar människors subjektiva förståelse av sin verklighet och det kollektiva meningsskapande som binder samman människor. Kunskap beskrivs som sociala konstruktioner, som uppkommer och får sin mening i samspelet mellan människor. Själva föreställningen om en objektiv, entydig verklighet är starkt ifrågasatt, både från filosofisk och sociologisk synvinkel. Ett uttryck för detta är att man talar om att ”fakta alltid är teoriladdade” (Alvesson och Sköldberg, 1994, s. 49), d v s de bygger på en mängd a priori-antaganden som inte går att belägga utan är en fråga om filosofiska och ideologiska ståndpunkter. Vi ”lägger alltså alltid in perspektiv i det vi ser, och inte nog med det, utan seendet är oskiljbart från perspektivet, är perspektiviskt” (ibid., s. 49). Utifrån detta synsätt har många vetenskapliga teorier med anspråk på att ge generella förklaringar kritiserats för att de tenderar att ensidigt och oreflekterat utgå från dominerande gruppers erfarenheter och intressen. Särskilt den feministiska forskningen arbetar med denna utgångspunkt, men liknande kritik av den etablerade vetenskapen har även formulerats utifrån klass- och post-koloniala perspektiv.

Utmärkande för postmodern eller poststrukturalistisk vetenskap är dessutom att den fokuserar språket och tecknen snarare än en yttre verklighet, som kan betraktas antingen som inte existerande, eller som oåtkomlig för vetenskapen. En annan typiskt postmodern eller postrukturalistisk ståndpunkt är förkastandet av ”de stora berättelserna”, d v s av teorier som gör anspråk på universell giltighet. Sådana teorier betraktas som

retoriskt präglade diskurser som egentligen bör ersättas med mikrohistorier, lokala, alltid provisoriska och begränsade berättelser som inte gör anspråk på att vara mer än de är, och som samtidigt inte borttränger de maktaspekter, friktioner, motsättningar och sprickor som oundvikligen uppträder i varje diskurs... (Alvesson och Sköldberg, ibid., s. 225).

Den metodologiska och epistemologiska utvecklingen inom vetenskapssamhället har, tillsammans med olika minoritetsgruppers växande kritik mot majoritetssamhällets normer för hur människor ska leva och tänka och vilken kunskap som är värdefull, också nått in i skoldebatten och läroplansdiskussionerna (t ex Gipps & Murphy, 1994; Hargreaves, 1995; Ålund, 1996 och Tesfahuney, 1999). När det gäller bedömningen av elevers kunskaper, så är ett resultat av detta ett ökat intresse för frågor som gäller validitet i vid bemärkelse (och en vidgning av validitetsbegreppet i sig). Exempel på sådana frågor är: Vad är det egentligen som proven testar och vilka "försanthållanden" om kunskap och mänskliga relationer bygger de på? Hur förhåller sig provens frågor till den specifika undervisningen? Gynnar formen eller valet av exempel systematiskt en eller annan grupp av elever? Vad beror det i så fall på?

Inte bara bedömningen, utan skolans innehåll och former i stort, har under de senaste åren, särskilt i USA och England, allt oftare diskuterats i förhållande till genus, etnicitet och i viss mån klass (t ex Banks, 1999 och Gipps & Murphy, 1994).

Ett annat perspektiv, som under samma period har varit mycket framträdande i diskussionerna och forskningen om bedömning sätter emellertid skolornas "effektivitet" i främsta rummet. Med "effektivitet" avses då den utsträckning i vilken arbetet i skolorna bidrar till det egna landets framtida ekonomiska utveckling. Som mått på skolornas effektivitet används elevernas betyg och resultat på storskaliga kunskapsprov. De amerikanska forskarna Madaus & O'Dwyer (1999) beskriver hur utfallet av sådana prov i USA och England används som underlag för att fatta beslut om allt från lönesättningen av enskilda lärare till nedläggning av skolor som inte "håller måttet".

Dessutom har läroplanerna över hela västvärlden under den senaste tioårsperioden gått mot en ökad fokusering av djupstrukturell kunskap, kognitiva färdigheter, reflexion, metakognition, etik och social kompetens (Hargreaves, 1999). Läroplanernas nya inriktning innebär att prov som är ensidigt inriktade mot återgivning av encyklopedisk kunskap inte längre är relevanta i förhållande till de officiella undervisningsmålen.

Den sociala ordningen i svenska klassrum

Tänkar om att undervisning ska utgå från elevers intressen och erfarenheter och vara individuellt anpassad, att förhållandet mellan elever och lärare ska vara personligt och innerligt, och att auktoritär disciplinering och själlöst pluggande är av ondo har uttryckts av många pedagogiska filosofer i historien, exempelvis Rousseau, Pestalozzi, Dewey och Key, och är alltså definitivt inte nya. Flera svenska författare (t ex Broady, 1982; Johansson, 1999; Lundgren 1979 och Rappe, 1973) har emellertid visat att sådana ideal knappast har varit utmärkande för verksamheten i det allmänna skolväsendet i Sverige. I den tidiga folkskolan, som i Lundgrens (ibid.) termer byggde på en moralisk läroplanskod, var ett centralt mål att inpränta fruktan för gud, fosterland och överhet hos eleverna. Läroverkens klassiska och senare folkskolans pragmatiska och realistiska läroplanskoder, var inte heller primärt inriktade på elevernas själsliga utveckling och behov och utgick inte heller från att lärare och elever skulle stå i förtroligt och personligt – eller jämställt! – förhållande till varandra.

Som Donald Broady gör en poäng av sin bok ”Den Dolda läroplanen” (1982), kom emellertid denna typ av värderingar i hög grad att präglade språket i läroplanen för den grundskola, som 1962 var avsedd att ersätta de tidigare skolformerna, inom det svenska skolväsendet. Bertil Gustafsson (1982) visar i sin bok med samma titel, att de följande två läroplanerna (LGR 69 och LGR 80) i än högre grad än denna första läroplan anbefaller elevcentrerade arbetssätt, individualiserad undervisning, intresse och aktning för elevens personlighet och psykologiska behov och systematiskt ersätter begrepp som kan associeras med en auktoritär utbildningssyn. Vidare poängteras att skolan inte får försöka utplåna de olikheter som elever bär med sig in i skolan, utan att var och en ska ha rätt att utveckla sin personlighet, samt att barn av naturen är nyfikna och kunskapsökande.

Slutsatsen som Gustafsson (ibid.) och hans kolleger drog av sina studier av vardagen i några svenska klassrum i slutet av 1970-talet, var dock att denna hade lite gemensamt med det ideal som postuleras i sådana formuleringar i läroplanen. Disciplineringen

av elevernas tankar och beteende var fortfarande genomgripande, även om den i vissa avseenden skedde mindre synligt. Däremot var det påtagligt att lärarna hade införlivat läroplanens begrepp och synsätt i sitt sätt att tolka och tala om sitt arbete. Donald Broady (1982) menar att iakttagelser som dessa – att skolan inte är vad den ger sig ut för att vara – är resultatet av en omöjlig ekvation som uppstår när en pedagogisk filosofi (här: progressivism och en romantisk elevsyn) som har sitt ursprung i och var praktisk i ett sammanhang (här: privatundervisning av enstaka elever i borgerliga hem), helt sonika ska överföras till ett helt annat sammanhang med helt andra förutsättningar (här: undervisning i klassrum av stora och socialt heterogena elevgrupper).

Att säga att disciplineringen inte har försvunnit ur den svenska skolan, utan tar sig andra uttryck, och att läroplanens värderingar kanske har haft störst betydelse på den diskursiva nivån, är dock inte detsamma som att säga att livet i svenska klassrum ser ut som för hundra år sedan – eller fyrtio. Naturligtvis ser mycket annorlunda ut. Graden av individualisering har säkerligen ökat – men i sådana former som uppfattas som möjliga och rimliga inom de diskursiva, praktiska och samhälleliga förutsättningar som präglar dagens skola. Individuella arbetsscheman och ”hastighetsindividualisering” (d v s varje barn tillåts arbeta i sin egen takt) är två vanliga slag av individualisering, som ger eleverna en viss frihet att disponera sin tid, och i vissa fall att välja mellan givna arbetsuppgifter. Att eleverna har något principiellt inflytande över innehållet eller uppgifterna är troligen inte lika vanligt.

Frykman (1998) beskriver hur elevernas fysiska rörelsefrihet under lektionerna var betydande i de klassrum som han har studerat under 1990-talet: eleverna har ofta lov att röra sig fritt i klassrummet för att hämta material eller ta hjälp av någon kamrat. Sällan krävs som förr total tystnad av eleverna när de arbetar, utan istället uppmuntras de ofta att hjälpa varandra att exempelvis lösa svåra uppgifter och att diskutera olika lösningsalternativ. Frykman menar, att kommunikationen mellan elever och lärare i svenska klassrum har blivit mindre ritualiserad och fått ett mera personligt innehåll. Exempelvis kallar eleverna ofta

sina lärare vid förnamn (istället för, som tidigare, ”fröken” eller ”magistern”) och det anses viktigt att lärare och elever känner varandra som ”hela personer” och inte bara utifrån den snäva elev – och lärarrollen. En annan sida av detta är att det är ”stötkigare” i klassrummen och att den auktoritet, som mera självklart förknippades med den gamla lärarrollen står under ständig förhandling. Spelet om makten över den sociala ordningen i klassrummet i ”progressivistiska” undervisningsmiljöer beskrivs väl av exempelvis Paul Willis (1981).

Har dessa förändringar när det gäller elevsyn och umgängeskod i klassrummet haft någon betydelse för hur man ser på och genomför bedömning? Ja, prov som, antingen de är av flervalformat eller består av essäfrågor, är inriktade mot återgivning av encyklopedisk kunskap, är ju inte förenliga med en strävan att eleverna ska uttrycka sin originalitet. Prov som används för att kontrollera i vilken utsträckning eleverna har tillägnat sig ett visst kunskapsstoff (och genom en poängsumma kommunicera om de har lyckats eller misslyckats) kan komma i konflikt med omtanken om elevernas psykiska välbefinnande – sådana prov orsakar negativ stress och försämrade självkänsla hos många elever, och riskerar leda till att elever som återkommande får låga resultat tappar all lust för skolarbete (se t ex Fransson, 1977 och Stiggins, 1998).

Inte heller är de traditionella provens inriktning mot *sortering* av eleverna förenlig med de demokratiska och humanistiska värderingar, som är viktiga grundprinciper i skolans självförståelse och som uttrycks i den offentliga retoriken om skolan. Enligt denna är ju skolans främsta uppgift att ge eleverna bildning¹¹ – inte att

¹¹ I läroplanskommitténs betänkande *En skola för bildning* (1992) betonas detta starkt. Enligt Broady, som där har skrivit det kapitel, som behandlar själva bildningsbegreppet, har detta genom tiderna används för att beteckna å ena sidan individens skapande av sig själv som självständigt tänkande och myndigt subjekt i världen, å andra sidan som ett visst gods av fasta kunskaper som förvaltas av samhällets sociala och kulturella elit. Broady pläderar för att undervisningen i det svenska skolsystemet ska vila på bildning i den första av dessa betydelser. Englund (1994) menar dock att denna bildningstanke så gott som försvunnit i de kursplaner som läroplanskommitténs betänkande var ägnat att lägga grunden till.

förbereda dem för olika positioner i det stratifierade arbetsliv, som de senare förväntas inträda i. I en bildningsskola motiveras inte eleverna av konkurrens (eller av tvång), utan av lusten att lära och upplevelsen av meningsfullhet. Det konkurrensmoment som finns inbyggt i de traditionella proven måste då ses som negativt eftersom det riskerar att ge eleverna en instrumentell syn på kunskap och en alienerad bild av sig själva som lärande.

Bedömningens konsekvenser

Den pedagogiska forskningen under 1990-talet har alltså uppmärksammat en rad negativa effekter av de "traditionella" provens negativa effekter på undervisning och lärande. Att externa, storskaliga, standardiserade prov påverkar lärarna att lägga upp sin undervisning med syftet att den egna klassen ska hävda sig väl, är något som ofta beskrivs i anglosaxisk litteratur (t ex Eckstein & Noah, 1993; Elliott, 1998, Gipps & Murphy, 1994, Taylor, 1994). Ett belysande talesätt för detta är: "the assessment tail wags the curriculum dog" (Wilson citerad i Nuttall, 1995). Stiggins (1999) refererar forskning som visar att denna typ av prov också i hög grad påverkar eleverna, på så sätt att de i genomsnitt ägnar en större andel tid åt att "plugga" för proven och att den tid de ägnar åt ämnen som inte testas på detta sätt blir mindre. Rapporter från amerikanska studier visar också att betygsgrundande storskaliga prov ger upphov till stress och nervositet hos många elever (Jones m fl , 1999).

Men vilka konsekvenser för elevernas lärande har sådana prov som utformas av de undervisande lärarna och som inte har lika omedelbara konsekvenser för elevernas betyg? Stiggins (1999) refererar en engelsk studie från 1998 av Black och William, som genom en metaanalys och syntes av 40 kontrollerade studier av förbättrad bedömningspraktik på klassrumsnivå, fann att dessa förbättringar resulterade i påtagligt bättre resultat på nationella prov. "Förbättringen" innebar framför allt att man utformat proven med syftet att ge lärare och elever maximal information om elevernas kunskaper, behov och progression. En annan slutsats var, att flera av studierna visade att de elever som dragit

störst nytta av dessa förändringar varde elever som vanligen presterade svagt. I dessa studier hade alltså förändringarna medfört att skillnaden mellan de högst och lägst presterande eleverna hade minskat.

Blacks (1999) slutsats efter en genomgång av flera empiriska studier av bedömning i klassrummet är, att en bedömningspraktik som hjälper eleverna att bli varse sin egen kunskapsutveckling och att lära sig hur man kan gå till väga för att åstadkomma arbeten som uppfyller kvalitetskriterierna inom olika ämnen, har större positiva effekter på elevernas motivation för skolarbetet än yttre belöningar. I klassrum där man tillämpar både en sådan bedömningspraktik och yttre belöningar, kommer den effekt som yttre belöningsystem har att dominera.

I en fallstudie gjord i ett svenskt högstadielklassrum (Korp, 1996), där man i SO-ämnena införde en typ av essäprov som syftade till att både mäta och befrämja förståelse och kritiskt tänkande, visade det sig att den förändrade provpraktiken fick störst positiva effekter på lärandet hos de elever som brukade prestera väl, medan lärandet i den lägst presterande gruppen snarare hade påverkats negativt. En slutsats som drogs i denna studie var att eleverna i klassen tolkade och hanterade provet och dess samband med undervisningen på olika sätt, och att sociala och kulturella faktorer hade betydelse för vilka förhållningssätt och inlärningsstrategier som olika elever använde. Men detta utesluter inte att förbättrade prov kan förbättra lärandet hos elever med olika social och kulturell bakgrund. Amerikanska studier (t ex Aness & Darling-Hammond, 1996 och Gordon & Bonilla-Bowman, 1996) och en engelsk (Black & William, 1998) visar att inlärningsstrategier och förhållningssätt till studier påverkas av undervisning och att olika former av bedömning, exempelvis "portföljer" och "självbedömning", faktiskt kan användas för att hjälpa elever att utveckla gångbara inlärningsstrategier.

Att elever riktar sin uppmärksamhet och sina ansträngningar mot vad de tror "kommer på provet", mer eller mindre medvetet anpassar sina inlärningsstrategier efter proven, och tar fasta på den kunskap som förefaller ekonomisk i förhållande till vad

som bedöms, är logiskt och har visats i både svenska och utländska studier (t ex Black, 1999 och Madaus & O'Dwyer, 1999). Unenge (1992), och Beach (1999) visar att det är svårt för lärare att motivera eleverna att tillägna sig andra kunskaper än dem som kontrolleras genom proven.

Internationella tendenser

Eckstein & Noah (1993) som har gjort en jämförande studie av åtta länders examinationssystem, drar slutsatsen att dessas utformning, syfte, funktioner och underliggande kunskapssyn kan härledas till respektive lands historiska och politiska utveckling. Som exempel har lokala examinationssystem tillåtits i länder där enskilda skolor och universitet har haft stort inflytande över utbildningens innehåll (t ex Tyskland), med motivet att detta skulle bidra till metodisk och innehållslig variation. I länder där man har försökt att konsolidera nationalstaten och via en starkt centraliserad utbildningspolitik överbygga regionala och religiösa motsättningar inom landet, har man istället betonat den sammanhållande funktionen hos examinationssystemet och eftersträvat nationell konformitet i bedömningen (Eckstein & Noah, 1993). Enligt Ball & Larsson (1989) har det allmänna skolväsendet i Sverige utmärkts av en relativt låg grad av central styrning, med stor variation i undervisningens innehåll och utformning som följd.

Gemensamt för flertalet västländer är emellertid att de under de senaste åren har ägnat mycket stora resurser åt att utvärdera och reformera sina nationella examinations- och betygssystem på alla nivåer. Den internationella "trenden" i detta sammanhang är, att man från statens sida har skärpt sin kontroll över utbildningen på systemnivå, medan ansvaret för bedömningen av enskilda elever i högre grad har lämnats åt de undervisande lärarna (Broadfoot, 1995). Lärarnas bedömningar har vidare blivit mindre formella och mera didaktiskt inriktade, och utvärderingarna på systemnivå mer standardiserade, tekniska och omfattande, enligt Broadfoot (ibid.). Distinktionen mellan "centraliserade" och "decentraliserade" utbildningssystem är därför mindre relevant.

I fråga både om nationella utvärderingar och nationella prov (exempelvis de svenska nationella proven och brittiska "A-level") har man emellertid gjort stora ansträngningar under 1990-talet för att göra frågorna mera utmanande och intressanta och för att "komma åt" mer sammansatt kunskap än vad som varit vanligt vid denna typ av prov. I USA har detta reformarbete inte enbart skett utifrån ett behov av att uppdatera bedömningspraktiken i förhållande till nya mål och metoder i den pedagogiska verksamheten, utan ofta varit avsett att utgöra själva drivkraften i en önskad förnyelse av skolans innehåll och arbetsformer (Ramirez, 1999).

Det ändrade förhållandet mellan bedömningen på individ- och systemnivå kan ses som en direkt konsekvens av utbildningens förändrade roll i samhället och för den enskildes livschanser. Utbildningssystemet måste ju hantera den frustration som uppstår i och med att allt fler individer, som egentligen skulle vilja lämna det så tidigt som möjligt, mer eller mindre tvingas till studier ända in i vuxenlivet¹². Samtidigt måste staten försäkra sig om att de enorma ekonomiska resurser som massutbildningen kostar används på effektivast möjliga vis. Att lärarna får ökat inflytande över bedömningen på individnivå, kan då ses som ett sätt att möta det första problemet – eftersom lärarna är de som både "känner" sina elever och planerar undervisningen, antas de vara de bäst skickade att utforma en bedömningspraktik som dels verkar motiverande för elever med svagt engagemang för sin utbildning, dels knyter an till den specifika undervisningen (Broadfoot, 1995).

¹² Jämför Per Anderssons och Bosse Bergstedts (1996, s. 107) problematisering av begreppet och fenomenet "livslångt lärande" som självklart relaterat till bildning och individers självförverkligande. Med hänvisning till Foucaults resonemang om disciplinering i skolor och korrektionsanstalter (i boken *Straff och övervakning*, menar de att: "man skulle kunna dra en parallell mellan inspärlandet av de vansinniga på institutioner och nutidens "inspärlande" av människor som annars inte skulle ha någon sysselsättning. (...) Följden kan bli att människor 'tvingas' kvar i utbildning, fastän inget hellre vill än att komma ut och arbeta."

En annan förändring som kan observeras i de flesta västländer, är att man, i den mån man har haft sådana, har frångått de normrelaterade (eller relativa) betygen och istället knutit betygs-sättningen till undervisningsmålen (d v s man har infört kriterierelaterade betyg). I Sverige skedde detta i samband med att skolan kommunaliserades och övergick från regel- till målstyrning. Betygen kom som ett mått på skolornas kvalitet då att göras till ett instrument för målstyrningen (Romhede, 1999). I samma reform ingick att läroplanens omfång minskades drastiskt och lärarna, i egenskap av professionella, tilldelades ansvaret för undervisningens form såväl som konkreta innehåll. Arbetsplaner med lokala profiler och mål och med lokala betygs-kriterier för varje ämne beslutas på de enskilda skolorna.

Vad innebär ”Det nya provparadigmet”?

De senaste par, tre decennierna har alltså medfört en rad förändringar som på olika sätt har påverkat förutsättningarna för sättet att se på bedömning i västvärlden. De nya synsätt på bedömning som därigenom har utvecklats har jag valt att kalla ”det nya provparadigmet”, men med det ska inte förstås att det nu finns ett enhetligt och tydligt perspektiv, som har ersatt ett annat. Vare sig det ”nya” eller det ”gamla” är enhetligt eller renodlat – dessutom existerar de parallellt. Det är snarare fråga om flera olika perspektiv och de bedömningsmetoder som kan hänföras till ”det nya provparadigmet” kan ha olika teoretiska och ideologiska grunder. Tydligt i forskningslitteraturen är ändå att diskussionen om bedömning under 1980-talet ändrar förtecken och får ett delvis nytt innehåll. Och även om dagens diskussion om bedömning är svårfångad och innehåller flera perspektiv, som skiljer sig åt på många punkter, så finns det alltså vissa gemensamma drag som har gjort det möjligt för forskare att tala om exempelvis en ”ny bedömningskultur” och ett ”nytt provparadigm”.

Exempelvis Broadfoot (1996) använder detta begrepp. Hon talar bl a om "en stark rörelse mot ett nytt provparadigm" (s. 15) och sätter startpunkten till slutet av 1980-talet. Birenbaum (1996) å sin sida menar att "en ny era" i kunskapsbedömningens historia började i mitten av 1980-talet och Calfee och Masuda (1997, s. 70) skriver att "bedömning på klassrumsnivå helt klart befinner sig i ett paradigmskifte". Förändringarna gäller vad, hur och varför man bedömer och vem som bedömer. Det "nya provparadigmet" utgör emellertid en integrerad del av vidare idéer om utbildning och kan, menar Shepard (2000), inte realiseras om inte "klassrumskulturen" ändras.

Shepard (ibid., s. 3) beskriver de några grundläggande didaktiska och teoretiska förändringar, som ligger till grund för "det nya provparadigmet":

Nya synsätt på utbildning

- alla elever kan lära sig
- utmanande innehåll inriktat mot komplext tänkande och problemlösning
- socialisering in i de vetenskapliga disciplinernas diskurser och praktiker
- autencitet i förhållandet mellan lärande inom och utanför skolan
- utveckling av strategier för tänkande och reflektion
- anslutning till demokratiska praktiker och trygg gemenskap

Kognitiva och konstruktivistiska inlärningsteorier

- Intellectuella förmågor är socialt och kulturellt utvecklade
- Elever konstruerar kunskap och förståelse i ett socialt sammanhang
- Nytt lärande formas av tidigare kunskap och kulturella perspektiv
- Intelligent tänkande involverar "metakognition" eller iakttagelse av det egna lärandet och tänkandet
- Djupförståelse är av principiell art och stödjer möjligheten att generalisera
- Kognitiva prestationer är beroende av situation och personlig identitet

Bedömning i klassrummet

- Utmanande uppgifter för att stimulera komplext tänkande
- Fokuserar inlärningsprocesser såväl som resultat
- En pågående process integrerad med undervisningen
- Används formativt för att stödja elevers lärande
- Förväntningarna är synliga för eleverna
- Eleverna är aktiva i utvärderingen av sitt eget arbete
- Används för att utvärdera undervisningen såväl som elevernas lärande

Begreppsliga ramar

”Det var enklare förr”

It used to be a simpler world, a world in which a test meant a number two pencil and an answer sheet. In such a world, everyone seemed to know what a test was and what it was for. Now, however, as the assessment world has expanded, there is considerable confusion about the purpose, role and interpretations of assessments. In the rush to develop more authentic assessments, to promote alternatives to traditional assessments, and to adapt assessment to the information needs of teachers, clarity of purpose has been lost. Chizek (1997, s. 8)

Hur prov kan se ut och kategoriseras, vad för slags kunskaper som de kan riktas mot, vad prov används till och vilka (ibland oavsiktliga) funktioner som de fyller och vilka teorier som är relevanta för att diskutera olika aspekter av bedömningsproblematiken, kommer att behandlas i de följande avsnitten. Hur man kan se på kvaliteten i kunskapsbedömning och hur proven påverkar undervisning och lärande kommer också att tas upp och, för att knyta an till den förra delens resonemang om förändringar som skett vad gäller kunskapsbedömningens roll i utbildningssystemet och samhället, diskuteras även olika sociala funktioner som bedömningen av elever eller studenter fyller. Slutligen görs ett försök att sätta de senaste decenniernas utveckling inom bedömningsområdet i relation till diskussioner om rättvisa och jämlikhet inom utbildningssystemet.

Teoretisk omorientering

”Det nya provparadigmet” innebär inte bara en ideologisk, utan också en teoretisk omorientering. Inträdet av nya synsätt och teorier om lärande som bas för den vetenskapliga diskussionen om bedömning utgör en viktig del av denna teoretiska omorientering, liksom förändringar i kunskapssynen. En annan är att

bedömningsforskningen nu delvis hämtar teorier från andra discipliner än förut.

Att nya discipliner har blivit aktuella för forskningen om kunskapsbedömning kan, menar Lundgren (1990), sättas i samband med den intresseförskjutning, som har ägt rum när det gäller bedömning – från form till innehåll och från urval av elever till urval av kunskap: ”Testteori, inlärningsteori, yrkessociologi och ekonomi blir mindre centrala för förståelsen av bedömning och utvärdering; epistemologi och etik, kunskapssociologi och utbildningshistoria blir relevanta” (ibid, s. 126).

Konstruktivism och kognitiva och sociokulturella teorier har blivit viktiga som förståelseram när det gäller lärande (t ex Gruber m fl , 1999), och detta har konsekvenser för sättet att se på hur man kan få kunskap om elevers lärande och vilka metoder som är lämpade för att kartlägga och beskriva elevers lärande eller kunskaper.

Vidare har kvalitativa ansatser till kunskapsbedömning gjort sig gällande i förhållande till de kvantitativa, naturvetenskapligt orienterade, som tidigare var normen (Eisner, 1991). Eisner (ibid,, s 176) skriver i en kritik av dessa: ”Naturvetenskapliga beskrivningar är ofta så distanserade, så formaliserade, så avskalade från de data, som de har härletts ur, att det är omöjligt för en läsare, att rekonstruera scenen: endast Gud vet vad det egentligen var som hände.” Eisner anser det ökade intresset för kvalitativ forskning vara ” en av århundradets viktigaste händelser” inom det pedagogiska området och påpekar att många framstående forskare inom pedagogiken har ändrat inriktning de senaste åren. Men också att många fortfarande håller fast vid de traditionella vetenskapliga normerna.

Men beror skillnader mellan olika forskares och praktikers inställning till bedömning alltid på att de har skilda filosofiska eller ideologiska utgångspunkter? Enligt Wolf (1994) är det inte så. Han poängterar att sådana skillnader mellan forskare eller mellan lärare likaväl kan bero på att de representerar olika ämnen, och att skillnader i synsätt ofta kan tillskrivas funda-

mentala olikheter mellan ämnenas natur, och vad man menar kan uppnås genom att studera dem. Wolf konstaterar utifrån en beskrivning, som Eisner gjort av humanistiska ämnen och hans argument för att studera sådana, att hans perspektiv skiljer sig markant från författarens som exempelvis Begle och Beberman, som är experter inom matematikutbildning. ”Skillnaderna kan inte avfärdas lätt”, skriver Wolf ”De är äkta och speglar olika inriktningar, som har sitt ursprung i respektive ämnes natur” (ibid., s. 2090).

Black (1999) poängterar, att varje bedömning av elevers kunskaper måste utformas som en konsekvens av hur man anser dels att lärande sker, dels vad som utgör progression i det aktuella ämnet: ”progress in learning physics (say) may be quite different from progress in learning history” (ibid., s. 123). Inte bara inlärningsteoretiska antaganden och allmänna utbildningsfilosofiska och epistemologiska ståndpunkter bestämmer alltså hur ”ett bra prov” ser ut, utan även själva ämnets epistemologi.

Vad som utgör relevant kunskap inom ett ämne är dock inte självklart, något som tydligt uttrycks av Kvale (1990, s. 121):

Kunskap inte bara är något givet, som kan testas och mätas genom ett prov; snarare är det så, att själva prövandet av kunskapen inom en disciplin innebär en operationell definition av disciplinens kunskap. På motsvarande sätt som den operationella definitionen av intelligens som det som mäts med ett intelligenstest, är en operationell definition av kunskap det som bedöms vid prov och examinationer.

Intresset för kunskapens sociala och historiska kontext genom-syrar många av de diskussioner och praktiska försök, som kan hänföras till ”det nya provparadigmet”.

Ett exempel på detta utgör denna tabell, ursprungligen sammanställd av Apple, här återgiven efter Gipps & Murphy (1994, s. 14).

FRÅGOR OM UNDERVISNINGEN	FRÅGOR OM BEDÖMNINGEN
Vems kunskaper utgör stoff i undervisningen?	Vilka kunskaper "räknas" i bedömningen? (are equated with achievement)
Varför har man valt dessa undervisningsformer för denna specifika elevgrupp?	Passar det sätt som elevernas kunskaper bedöms på (och bedömningens innehållsliga fokus) för olika grupper och individer?
Hur kan man undervisa om kvinnors och olika etniska gruppers historia och kultur på ett sätt som gör dem rättvisa?	Återspeglas vidden av kulturella erfarenheter i kunskapssynen och bedömningskriterierna? Vilken betydelse har individers skilda kulturella erfarenheter för deras preferenser när det gäller att kommunicera sina kunskaper och vilka konsekvenser kan det få för värderingen av deras prestationer?

TABELL 1. Frågor om undervisning och bedömning i relation till jämlikhet

Inlärningsteorier

Då prov i allmänhet är konstruerade med avsikten att mäta effekter av elevers lärande, är teorier som på ett eller annat sätt rör lärande centrala i diskussioner om bedömning. Sådana teorier bygger på olika definitioner lärande och kunskap, och innefattar hypoteser bl a om hur lärande går till, vilka förutsättningar som är gynnsamma för lärande, och hur effekter av lärande kan observeras. Tre viktiga inriktningar inom inlärningsteorin är (1) behaviourismen (2) kognitiv och metakognitiv teori och (3) sociokulturella inlärningsteorier. De två senare delar många väsentliga drag och skiljer sig båda markant från behaviouristiska teorier, främst i det att de ser lärande som individens aktiva skapande av mening – de vilar alltså på en konstruktivistisk kunskapsyn. Det finns också väsentliga skillnader mellan kognitiva och sociokulturella teorier. De sociokulturella teorierna har ofta utvecklats ur de kognitiva utifrån en kritik av dessas, som man har menat, ensidiga fokusering av individen och vad som händer ”inuti huvudet”. I det sociokulturella perspektivet ses individens meningsskapande som beroende av de sociala och kulturella sammanhang som den lärande är en del av.

Medan behaviouristiska inlärningsteorier utgår från en perception av intelligens som endimensionell, medfödd, konstant och normalfördelad, så bygger de kognitiva och sociokulturella teorierna på nyare uppfattningar av intelligens, som sammansatt av flera dimensioner och som resultatet av en kombination av biologiska förutsättningar och erfarenheter som individen gör i interaktion med den sociala och fysiska omvärlden. Enligt denna nyare uppfattning är intelligensen alltså påverkbar och beroende av exempelvis sociala, kulturella och materiella villkor och av undervisning (se t ex Wolf & Reardon, 1996 och Berhanu, 2002).

Behaviourismen

I den behaviouristiska traditionen betraktas lärande som en förändring i en persons kunskap eller "kognitiva associationer", som leder till ett förändrat beteende. Lärandet manifesterar sig alltså i observerbara beteendeförändringar (Cizek, 1997). Vidare ses lärande som linjärt och sekventiellt – komplex förståelse kan endast uppnås genom tillämpning av elementära kunskaper, som lärts in tidigare (ibid.)¹³. På behaviouristisk inlärningsteori byggs sådan undervisning, som utgår från en bild av eleven som ett tomt kärl, som ska fyllas, eller en tom tavla, "tabula rasa", där läraren förmedlar kunskap till eleven, vars enda uppgift är att lagra den (Birenbaum, 1996). I den behavioristiska perceptionen spelar alltså minnet och förmågan att reproducera en central roll. Birenbaum citerar Freire: "Ju mer fullständigt han [läraren] fyller kärlet, desto bättre lärare är han. Ju mer beredvilligt kärlet låter sig fyllas, desto bättre elever är de" (ibid, s. 4). Freire benämner detta synsätt "the banking concept".

Fram till slutet av 1970-talet var behaviourismen den totalt dominerande inlärningsteorin och, enligt Eisner (1991), har dess tidiga företrädare Skinner och Thorndike haft stort inflytande när det gäller sättet att se på kunskapsbedömning. En ambition som drev Thorndikes forskning var, skriver Eisner, att utveckla en vetenskapligt utformad utbildning, som med hjälp av vetenskaplig utvärdering av elevernas kunskapsstilläggande, skulle ge optimal avkastning på de statliga utbildningsinvesteringarna. I slutet av 70-talet började emellertid den behaviouristiska hegemonin att utmanas. Glaser & Silver (1994) ser det som en följd av att de beteendebaserade teorierna inte är användbara för att beskriva de komplexa kognitiva processer, som är involverade i problemlösning och reflektion, som kom att spela en allt större roll i undervisningen och röna ett ökat intresse bland både lärare och forskare. Nya, kognitiva teorier var därför nödvändiga.

¹³ Jämför med Bernsteins term "samlingskod", som innebär, att eleven måste lära sig fakta före sammanhang, och från "kunskapens ytstruktur" successivt närma sig dess djupstruktur, som endast kan bli tillgänglig för fåtalet.

Kognitiva och metakognitiva teorier

Piaget menade att vi lär genom att handla, genom att vi självständigt försöker att lösa problem och på så sätt ta kontroll över världen. Han menade också att abstrakt tänkande har sitt ursprung i konkret handlande (Black, 1999). Dessa båda principer är grundläggande för kognitiva teorier om lärande.

Om Piagets teorier är begränsade till utvecklingen av logiskt och abstrakt tänkande, så utgår kognitiv teori emellertid numera, enligt Black (ibid.), från att också känslor och attityder har betydelse för tänkandet. Man söker därför inte efter en allmän förmåga till abstrakta formella tankeoperationer hos individer, utan utgår från att individers kognitiva kapacitet i viss utsträckning varierar från situation till situation. Signifikativt för ett sådant vidgat kognitivt perspektiv är Wittrocks definition av lärande som

”de processer, som är inblandade när en person förändras genom sina erfarenheter; en process, som resulterar i relativt permanenta förändringar i förståelse, attityder, kunskap, information, förmåga och färdighet genom erfarenhet. (Wittrock i Cizek, 1997, s. 3)

Cizek (ibid, s. 134) återger också Good & Brophys kommentar till denna definition:

Definierat på detta vis kan inte lärande likställas med observerbara prestationer. Det är sant att lärandet förändrar prestationskapaciteten, och att vi måste observera hur prestationerna förändras för att kunna dra slutsatser om huruvida lärande har ägt rum (...) Likväl är den prestationspotential, som har utvecklats genom lärande, inte desamma som dess reproduktion eller tillämpning i någon specifik bedömningsituation.

Med andra ord: att en elevs prestation inte indikerar att lärande har ägt rum, utesluter inte att eleven har lärt sig! Exempelvis mognande, eller utveckling kan inte i sig betraktas som lärande, då lärande är en process som inbegriper både individens erfärande och ett aktiv hanterande av information och konstruktion av kunskap, enligt Cizek (ibid.).

Enligt t ex Marton (1997) kan lärande definieras som att det sker en kvalitativ förändring i det sätt på vilket någon uppfattar något. När man lär sig är det alltså inte i första hand fråga om en kvantitativ förändring – att man kan mer än innan, utan om att man förstår ett fenomen på ett nytt sätt.

Lärande i denna mening är liktydigt med vad Piaget betecknar som "ackommodation". Med ackommodation avses den process där en individ ändrar sina kognitiva scheman. Detta sker när individen erfar något som inte kan göras begripligt i relation till befintliga kognitiva scheman – d v s när hennes eller hans tidigare föreställningar utmanas av nya fakta. Undervisning som bygger på kognitiv teori tar därför sin utgångspunkt i elevernas tidigare föreställningar om aktuella fenomen, och lärarens uppgift är att försätta eleverna i situationer där de kan göra erfarenheter som ger upphov till kognitiva konflikter (Black, 1999), men också "erbjuda tillfällen för eleverna att använda det de redan kan och vet för att förstå nytt material" (Birenbaum, 1993, s. 6).

Undervisning, som utgår från kognitiv inlärningsteori, kallas ibland kognitiv pedagogik, och utmärkande för sådan pedagogik är, att elever istället för att betraktas som tomma tavlor eller kärl, som ska "fyllas på" med utifrån kommande kunskap, så ses de som "noviser" eller "lärjungar", som av naturen är aktivt sökande och vars lärande gagnas av att de deltar i målorienterade, kollaborativa aktiviteter" (Reid & Stone, i Birenbaum, 1996, s. 8). I formuleringar som denna blir det tydligt att kognitiv och konstruktivistisk inlärningsteori i vissa avseenden är nära besläktade med sociokulturella teorier om lärande.

Eftersom den lärande i kognitiv teori ses som aktiv skapare av sin kunskap, och lärandet som en process som utgår från förförståelsen, så är det ur detta perspektiv viktigt att eleven utvecklar effektiva strategier för sitt lärande, d v s "grundläggande metoder för att behandla och använda information" (Phye, 1997, s. 106). Sådana metoder benämns av Phye "kognitiva strategier". Phye nämner fyra kognitiva strategier:

1. *Analys* – att sönderdela en helhet i sina beståndsdelar och förstå förhållandet mellan delarna från ett helhetsperspektiv.
2. *Jämförelse* – att identifiera likheter och skillnader och förstå betydelsen av dessa ur ett större perspektiv
3. *Slutledning och tolkning* – att använda olika former av induktivt och deduktivt resonerande, inklusive tillämpning av regler och principer, för att komma fram till en slutsats eller lösa ett problem.
4. *Värdering* – att avgöra vad man ska tro eller göra baserat på explicita kriterier och fakta.

Med ett kognitivt perspektiv på lärande, är det inte bara viktigt att eleverna utvecklar fungerande kognitiva strategier, utan ett mål för undervisningen bör också vara att eleven ska lära sig om sitt eget lärande, d v s de ska kunna reflektera över, utvärdera och förbättra sina kognitiva strategier, sitt tänkande. Detta benämns av bl a Phye som ”metakognition”, och de ”metakognitiva strategier”, som undervisningen bör syfta till att eleverna ska utveckla är, enligt honom: 1. *Planering*, som innebär att problemet identifieras, att man jämför dess komponenter med problem som man mött tidigare och att man försöker finna lämpliga strategier för att gripa sig an problemet. *Utkast och utprovning*, som innebär att man provar ett antal angreppssätt för att försöka lösa uppgiften *Kontroll och revision*, som innebär att man kontinuerligt kontrollerar om de strategier man tillämpar är framgångsrika, om man uppfyller delmål och närmar sig en lösning. 4. *Utvärdering och reflexion*, där man ser tillbaka på den process som ledde fram till resultatet och utvärderar de metoder man använde, samt reflekterar över den egna effektiviteten.

Metakognition beskrivs ofta som ett medel för att uppnå ”självreglering” (self regulation), dvs att eleverna ska bli självständiga och kunna styra sitt eget lärande. Med hjälp av adekvat undervisning kan elever förbättra sin metakognition. Black (1999) nämner flera studier, som påvisar samband mellan graden av metakognition hos elever och deras studieresultat. Exempelvis Brown & Ferrara (1985) har visat att elever som har en historia av svaga prestationer i skolan och som har fått hjälp med att utveckla sin metakognition har gjort snabba framsteg.

Exempelvis Wolf & Reardon (1996) pekar på hur denna typ av rön och synen på intelligens som utvecklingsbar innebär en möjlighet att utmana den traditionella, dominerande, med "den gamla" psykometrin befästa föreställningen att "excellence" och jämlikhet står i ett motsatsförhållande. Gordon och Bonilla Bowman (1996) menar att bedömningen av elevernas kunskaper kan fylla en viktig funktion för utvecklingen av deras kognitiva och metakognitiva strategier, eftersom de dels kan göra eleverna uppmärksamma på sin egen kunskapsprogression, dels kan hjälpa dem att förstå vad dessa författare kallar "the anatomy of good work", dvs vad som karaktäriserar arbeten av hög kvalitet.

Sociokulturella perspektiv

Den viktigaste gestalten när det gäller sociokulturella teorier om lärande är naturligtvis Lev Vygotsky – grundaren till den ryska kulturhistoriska skolan. Vygotsky ansåg att lärarens uppgift ska vara att identifiera var eleven kunskapsmässigt befinner sig och hur hon eller han skulle kunna förmås att utvecklas från detta utgångsläge med rätt stöd av omgivningen, och sedan att hjälpa eleven att nå denna potential. Utrymmet mellan vad någon kan göra utan hjälp, respektive med hjälp från andra, kallar Vygotsky för "proximala utvecklingszoner". Läraren hjälper eleverna att röra sig ut mot, och flytta fram, gränserna för sina proximala utvecklingszoner genom att ge uppgifter, som varken är triviala eller alltför krävande (Black, 1999).

Så långt liknar den lärarroll som föreslås av Vygotsky lärarrollen i den kognitiva pedagogiken. Sociokulturella teorier om lärande har också, som redan nämnts, mycket gemensamt med de kognitiva – framför allt sedan också dessa har börjat omfatta attityder och känslomässiga aspekter på lärandet och betrakta kapaciteten för abstrakt tänkande som situationsbundet. Forskare som Lauren Resnick beskrivs också omväxlande som representanter för kognitiva och sociokulturella teorier. Men de kognitiva teorierna om lärande har också kritiserats av forskare med sociokulturell inriktning för att vara "all in the head" (Chizek, s. 7) och för att isolera lärandets kognitiva aspekter från de sociala (Claesson, 1999).

Utmärkande för det sociokulturella perspektivet är just intresset för lärandets sociala karaktär och sammanhang. Man talar om "situerat lärande" och "situerad kognition", begrepp som riktar uppmärksamheten mot att lärande alltid sker i en bestämd situation och att de resulterande kunskaperna är förbundna med den lärandes uppfattning av denna situation. I sociokulturella teorier om lärande läggs stor vikt vid den språkliga kommunikationen och vad den betyder för olika personer i olika sammanhang (i forskning om situerat lärande fokuseras emellertid inte nödvändigtvis den verbala kommunikationen). Huruvida undervisning är att betrakta som effektiv är ur detta perspektiv därför till stor del en fråga om i vilken utsträckning dess språkliga interaktionsmönster delas av eleverna., menar t ex Black (1999).

Jean Lave, Lauren Resnick och Barbara Rogoff är några av de internationellt mest aktuella företrädarna för ett sociokulturellt perspektiv på lärande. Lave, som är antropolog har intresserat sig för sådant lärande som sker i vardagslivet utanför formella utbildningsinstitutioner. En av hennes teser är, att man måste överge synen på kunskap som något entydigt och universellt, som individer bär med sig och tillämpar i olika situationer. Gruber m fl (1999, s. 217), skriver angående Laves uppfattning om "kognitionens kontinuitet" (d v s i vilken utsträckning tankeoperationer av olika slag – exempelvis matematiska – kan anses vara desamma i olika situationer), att hon ser det som en fråga om social reproduktion, snarare än som en kognitiv akt. Tillsammans med Etienne Wenger har introducerade Lave begreppet "legitimt perifert deltagande" (Lave & Wenger, 1991), utifrån en studie av bl a engelska slaktar- och skräddarlärlingars och afrikanska barnmorskeelevers lärande på de arbetsplatser där de praktiserade. Deras slutsats i denna studie var att dessa yrkes elever tillägnade sig sin yrkeskompetens genom att delta i den sociala praktik, som arbetsplatsen utgjorde och i vilken de i rollen som lärling kunde delta först perifert, och till slut som fullvärdiga medlemmar. Lärandets kognitiva dimensioner kunde i denna process inte skiljas från de lärandes identitetsutveckling (ibid.).

Lauren Resnick har bl a studerat hur barn läser och räknar i och utanför skolan och funnit stora skillnader – eleverna hanterar exempelvis matematiska problem helt annorlunda i skolan jämfört med hur de gör och tänker när de ställs inför motsvarande problem i vardagen. Resnick menar, att för att man ska kunna dra några slutsatser om en elevs kognitiva aktivitet, så måste man ha en bild av elevens subjektiva konstruktion av den situation där aktiviteten äger rum (Gruber m fl , 1999). Kunskaper betraktas i sociokulturella teorier som sociala konstruktioner och som sådana formade av historiska villkor – sociala, politiska, kulturella. Resnick pekar på att också de sätt på vilka olika personer ser på kunskap och lär sig varierar mellan olika tider, kulturer och samhällsklasser. Inte bara kunskapen, utan även våra kognitiva, och metakognitiva strukturer är alltså socialt och kulturellt betingade, menar Resnick (ibid.).

Barbara Rogoff delar denna Resnicks uppfattning, att människors tänkande till innehåll såväl som form är bestämda av de sociala och kulturella sammanhang där de har sitt upphov. Rogoff är särskilt intresserad av barndomen och den tidiga socialisationen inom familjen, som hon menar är mycket viktig för individens kognition senare i livet (Gruber m fl, 1999). Med utgångspunkt i Vygotskys teori om proximala utvecklingszoner, betraktar hon barnets lärande som något som sker i ett samspel mellan barn och vuxna, där de tillsammans löser uppgifter som de har gemensamt intresse av. Rogoff ser barnet i sådana situationer som en sorts "lärling" (jämför Lave ovan), som genom att identifiera sig med den vuxne gradvis tillägnar sig dennes kognitiva mönster (ibid.).

I Sverige är Roger Säljö en stark företrädare för sociokulturell teoribildning kring lärande. I Säljös forskning fokuseras språkets betydelse för barns kognitiva utveckling och språket som bärare av kognitiva, sociala och kulturella strukturer (Säljö, 1994 och 2000).

Vad får det sociokulturella perspektivet för konsekvenser när det gäller hur man ser på bedömning? Vygotskys resonemang om proximala utvecklingszoner, vilket av Gipps (1994) uttrycks

som att det ur undervisningssynpunkt är mera intressant att veta vad en elev kan klara av att göra under optimala förhållanden, än vilken som är hans eller hennes mest typiska prestation, får som följd att det är rimligare att läraren i en provsituation hjälper eleven att prestera sitt bästa, än undanhåller hjälp för att eleven ska prestera sitt mest typiska (ibid.). En konsekvens av uppfattningen att elevers förståelse av språket, och sätt att själv använda språket, är beroende av relationen mellan de språkliga koderna i uppväxtmiljön och i skolan, är att man måste var mycket noggrann, både med provets språkliga utformning och med bedömningsituationen som sådan. En annan konsekvens är att det är viktigt att variera provformerna och en tredje gäller vilka slutsatser man drar om individer och grupper på basis av resultat från prov och tester, exempelvis intelligenstest. Att tidigare mätningar har visat på att intelligensen är normalfördelad inom befolkningen och att den genomsnittliga intelligenskvoten skiljer sig åt mellan grupper, säger ur detta perspektiv inte så mycket den intellektuella kapaciteten i olika grupper. Utfallet kan lika gärna förklaras med att den operationella definitionen av intelligens, d v s att intelligenstestet inte är neutralt i förhållande till olika grupper, och att olika grupper har haft olika tillgång till metakognitiv träning, exempelvis genom utbildning (Gipps & Murphy, 1994).

Hur kan man kategorisera prov?

I litteraturen används olika indelningsgrunder som utgångspunkt för beskrivningar av olika slags bedömningar. Några av de vanligaste är:

- A. *Nivån för sammanställning av resultaten:* Vem gäller bedömningen? Ska resultatet beteckna kvaliteten hos exempelvis en enskild students prestation, eller för en grupp studenter, eller ska det gälla för en hel skola, eller för en nation?
- B. *Bedömningens betydelse.* Är bedömningens resultat avsett att ligga till grund för beslut som är av stor betydelse antingen för enskilda individer eller för t ex en skola eller utbildnings-

systemet, eller ingår den i ett mindre kritiskt sammanhang?
Hur viktig är likvärdigheten?

- C. *Bedömningens syfte.* Görs bedömningen i syfte att generera ett omdöme eller ett betyg, dvs att förmedla resultatet av en värdering av det som studeras (t ex en elevs kunskaper inom ett visst ämne), eller är avsikten med bedömningen i första hand att bidra till utveckling och förbättring?
- D. *Bedömningens objekt.* Vad fokuseras i bedömningen? – exempelvis faktakunskaper, problemlösningsförmåga, estetisk förmåga, yrkesskicklighet? Eller gäller den kvaliteten i verksamheten? Gäller bedömningen resultatet eller arbetsprocessen?
- E. *Bedömningsmetod.* Vad för slags instrument använder sig exempelvis en lärare av för att ta reda på vad en elev kan, vet eller har lärt sig? – exempelvis informella observationer, skriftliga prov, eller praktiska uppgifter.
- F. *Bedömningsprincip.* I relation till vad görs värderingen? Ska värdet av exempelvis en persons prestation bestämmas i relation till hur övriga i gruppen presterat, eller i förhållande till på förhand uppsatta kriterier? Eller ska bedömningen ske mera förutsättningslöst, och resultera i en kvalitativ beskrivning?
- G. *Vem som utformar respektive genomför bedömningen.* Har provet utformats centralt eller lokalt, har det konstruerats av en undervisande lärare, en läromedelsförfattare eller genom statlig försorg? Och om bedömningen gäller exempelvis en skola eller ett utbildningssystem – är det fråga om en extern eller intern bedömning? Är det den undervisande läraren som bedömer sina elever eller bedömer eleverna sig själva (självbedömning)? Bedömer eleverna varandra (kamratbedömning)?
- H. *Beslutsfattare:* Vem ska använda resultatet av bedömningen? – exempelvis undervisande lärare, elever eller studenter, skol- eller utbildningsledning, eller så kallade "policy-makers".

Bedömning på individ- och systemnivå

Begreppsapparaten kring kunskapsbedömning är invecklad och mångtydig – det gäller såväl inom svenskan som engelskan. Begrepp som i ett sammanhang används för att beteckna en viss fas i bedömningsprocessen, kan i ett annat sammanhang användas som beteckning för ett specifikt filosofiskt eller pedagogiskt synsätt kring bedömning. Något som komplicerar relationen mellan begreppen är också, att ”assessment” i brittisk litteratur ibland även används som en övergripande beteckning för bedömning på såväl system- och gruppnivå som individnivå, d v s synonymt med det svenska begreppet bedömning. På samma sätt är det utanför kretsen av utbildnings- och bedömningsforskare, där termen ”assessment” kan referera till all slags värderande, bedömande och utvärderande verksamhet (exempelvis medicinsk diagnos). Det förekommer också, främst hos amerikanska författare, att ordet ”evaluation” används med denna generella innebörd.

Dessutom har de svenska begreppen bedömning och utvärdering och de engelska begreppen assessment och evaluation alla det gemensamt, att de kan avse (1) företeelsen i allmänhet (2) de konkreta instrument, som används för att skapa information om det som ska bedömas (d v s prov, utvärderingsformulär, arbetssuppgifter etc.), (3) själva den process, där denna information tolkas, värderas och ”betygsätts”. Det sista sammanfaller med Lauvås (1998) beskrivning av det norska begreppet ”evaluering”, som ”å sammenholde en beskrivelse karakteristikk av ett fenomen med kriterier (normer, standarder, værdare) for att komme fram til konklusjon(er)” (s. 15).

I detta avsnitt ska jag, utan att fördjupa mig i detaljer mer än jag finner nödvändigt, föreslå hur några vanliga begrepp inom området kan förstås och användas för att analysera och diskutera egen eller andras bedömningspraktik ur olika aspekter.

Det svenska ordet ”bedömning” kan användas för att beteckna utvärderande verksamhet riktade mot både individer och skolor, men också abstrakta enheter som undervisning och läroplaner. Harris och Bell (1994) skiljer mellan bedömning som fokuserar *de lärande* respektive bedömning med fokus på *de resurser som mobiliseras för att främja lärande* – exempelvis böcker, undervisning och lärare. Det senare refereras oftast till som *utvärdering*.

Att det är systemnivån som fokuseras och inte enskilda individer, hindrar dock inte att utvärderingar ofta kan basera sig på bedömningar av individuella prestationer. Så är t ex fallet med de svenska sk nationella utvärderingarna, som genomförs på Skolverkets uppdrag med det huvudsakliga syftet att kontrollera att kommunerna fullföljer sitt uppdrag att ge grundskoleutbildning som motsvarar läroplanens krav. Andra utvärderingar kan basera sig på elevernas betyg, men denna information sammanställs och analyseras då på gruppnivå. Det innebär att man inte fäster något avseende vid vilka individer som representerar olika poängsummor, utan resultatet används för att dra slutsatser om skillnader mellan exempelvis flickor och pojkar, olika skolor, svenska elever och elever med utländsk bakgrund osv, om kunskapsstandarden i olika ämnen, eller om förändringar i kunskapsstandarden. Förutom på betyg och provresultat kan utvärderingar också bygga på data av helt andra slag (t ex observationer, enkäter och olika kvalitetsindikatorer såsom personella resurser, lokaler, utbildningsnivå etc.)¹⁴, samt på en kombination av data från olika källor.

Enligt ovanstående resonemang kan alltså utvärdering ses som en särskild typ av bedömning. Men begreppen används ibland också ”tvärtom”: Enligt Michael Scriven, USA, som är en förgrundsgestalt inom utvärderingsforskningen, betecknar begreppet utvärdering dels processen att fastställa meriten, kvaliteten eller värdet hos olika ”enheter”, dels resultatet av en sådan process. Några ”enheter”, som Scriven nämner som aktuella att

¹⁴ Jämför t ex ECERS, det instrument som används internationellt för att utvärdera kvaliteten i barnomsorgen.

utvärdera, är program, produkter, personal, styrdokument och prestationer. Med denna definition kan bedömning av enskilda individers kunskaper ses som ett specialfall av utvärdering. Oftast används begreppet dock för att beteckna bedömning på systemnivå (se t ex Choppin, 1990; Worthen & Van Dusen, 1991). Även om samma begrepp och teorier i princip är giltiga för bedömning av individer, finns det de (exempelvis Choppin, ibid.) som menar att termen faktiskt bör reserveras för sådana "abstrakta enheter" som utbildningssystem, program och läroplansmål.

Utvärderingar och utvärderingsforskning ska dock inte ges någon mer specifik behandling i denna skrift, eftersom ju fokus ska ligga på "vanliga prov", d v s bedömning som sker i klassrummet av enskilda elevers kunskaper.

Bedömning av individers kunskaper

När det gäller bedömning som görs i syfte att dra slutsatser om individers kunskapsutveckling eller kompetens, finns än mindre konsensus om vilken term som ska användas. På norska använder Per Lauvås (1998) termen "individevaluering", och han menar att det svenska ordet "bedömning" just är att betrakta som en "summativ individevaluering" (s. 15). Denna starka avgränsning ser jag emellertid som mindre lämplig av det skälet att den ju lämnar det svenska språket utan begrepp för bedömning i mera allmän bemärkelse. "Individbedömning" är den term som jag skulle vilja föreslå används på svenska, eftersom den samtidigt som den är tillräckligt tydlig, inte låser fast betydelsen till något särskilt syfte eller synsätt.

"Mätning" eller "testning" av elevers kunskaper är en särskild form av bedömning. Enligt Choppin (1990) är pedagogisk mätning framför allt att betrakta som en grundläggande forskningsprocedur. Därför har, anser han, "mätning" egentligen inte någon direkt roll vare sig i det pedagogiska arbetet i klassrummet eller för registreringen av elevernas kunskaper och progres-

sion. De flesta utbildningssystem i världen registrerar visserligen elevernas kunskaper och prestationer på något sätt, men, påpekar Choppin, betygen baserar sig på en informell syntes av data från många olika källor, och kan inte uppfylla de krav på precision och objektivitet som är förknippade med vetenskaplig mätning.

Pedagogisk mätning i Choppins mening benämns ofta "testning" och kan gälla såväl attityder som förvärvade kunskaper och färdigheter och inneboende förmågor av olika slag, exempelvis spatial eller empatisk förmåga, intelligens, läsförmåga eller ordförråd. Flera av dessa variabler kan inte mätas direkt (de beskrivs som "latenta"), utan man får dra slutsatser om dem på grundval av testtagarnas resultat på "manifesta variabler" – dvs antal rätta svar på frågorna i testets olika delar (numera avläses de latenta variablerna i statistiska korrelationer mellan de manifesta)¹⁵.

Testinstrumenten, vilka helt enkelt benämns som "test", består oftast av ett frågebatteri med tillhörande svarsalternativ. Testerna är externt utformade och vetenskapligt utprovade och bygger på antagandet att i den mån som de yttre betingelserna hålls under kontroll och är identiska för dem som gör testet, så kommer resultaten att ge en sann bild av den förmåga, attityd eller liknande, som studeras. Testresultaten kan sammanställas antingen på individ- eller gruppnivå, beroende på testningens syfte. I skolan används intelligens- och olika psykologiska tester av psykologer för att diagnostisera elever med exempelvis misstänkta begåvningshandikapp eller neurologiska skador. Tester används också inom forskningen om bl a individuella differenser och för att utveckla den vetenskapliga förståelsen av olika typer av intelligens. Psykologiska tester och bedömning i forsknings-syfte faller dock utanför ramarna för vad som diskuteras i denna text. Värt att notera är emellertid, att i engelskan betecknar ordet "test" inte bara psykometriska bedömningar, utan alla

¹⁵ Detta har blivit möjligt sedan införandet av så kallad multivariat analys, en avancerad statistisk behandling, som görs med hjälp av dataprogram som MANOVA och LISREL (Lohman, 1997, Rosén, 1999).

möjliga slags prov även ”vanliga prov” av det slag som utformas av de undervisande lärarna och är en del av det dagliga klassrumsarbetet.

Enligt Worthen & Van Dusen (1991), handlar ”mätning” om att på grundval av empiriska observationer göra en ”enkel kvantitativ beskrivning” av det som studeras. Vid mätningen görs ingen värdering av det som mäts eller av resultatet av mätningen. Så säger t ex resultatet av intelligensmätning ingenting om meritvärdet av intelligens i olika sammanhang (ibid.).

Bedömningens betydelse

Hur viktig utgången av en bedömning är för den som det gäller varierar naturligtvis. Ibland kan mycket stå på spel för den enskilde, som t ex vid högskoleprovet, eller för andra aktörer, som t ex på den internationella valutamarknaden, där länders ekonomier är beroende av utländska börsanalytikerns bedömningar. I engelskan används ofta termerna *high stakes* och *low stakes* för att beteckna ”ändpunkterna” på skalan mellan bedömningar som är avsedda att få konkreta och betydelsefulla konsekvenser (t ex för individens livschanser eller för huruvida man ska genomföra en utbildningsreform) och dem som inte är det. Körkortsprov och sådana utvärderingar av högskolor som görs med syftet att besluta om examensrättigheter är exempel på bedömningar som tillhör den första kategorin. Den andra kategorin omfattar sådana bedömningar, som inte är avsedda att ligga till grund för denna typ av kritiska beslut och där resultatet ofta stannar mellan t ex lärare och elever för att användas som en del i det fortsatta pedagogiska arbetet.¹⁶

¹⁶ När det gäller bedömning av elever, så visar dock t ex Torrance (1999) att även de mest vardagliga bedömningsituationer (t ex muntliga läxförhör) har viktiga konsekvenser för elevers intellektuella självförtroende om de ingår i ett mönster av bedömningsituationer, som sammantaget kommunicerar en tydlig bild till eleven om dennes förmåga. Till synes betydelselösa bedömningar kan på så sätt påverka såväl motivation och resultat som studieval hos individer och därmed faktiskt vara av stor betydelse för deras framtid.

Bedömningens syfte

Vilken typ av slutsatser är man intresserad av att kunna dra på basis av bedömningsresultaten, och varför? Grovt kan man skilja mellan sådana bedömningar som primärt är avsedda att resultera i ett omdöme om det som studeras och sådana som syftar till att generera sådan information, som kan användas för att förbättra det som studerats, exempelvis elevers lärande och lärares undervisning. Michael Scriven introducerade 1967¹⁷ termerna ”summativ” och ”formativ” för att beteckna de två typerna. I en senare text (Scriven, 1991, s. 17) demonstrerade han skillnaden mellan dem med ett citat av Robert Stake: ”Formativ bedömning är när kocken smakar soppan och summativ bedömning är när gästen gör det”

Summativa bedömningar inbegriper alltid ett värderande moment. Men också formativa bedömningar kan innehålla utlåtanden om kvaliteten – det som skiljer den formativa bedömningen av elevers kunskaper från den summativa är, att den har ett pedagogiskt syfte och inte används för selektion eller rangordning (ibid.). I den ursprungliga meningen refererar termerna till ”skillnaden mellan utvärderingar som syftar till att förbättra något (t ex ett program) och utvärderingar för alla andra syften (framförallt för att fatta beslut och dra slutsatser om det)” (ibid., s. 18). Sedan termerna ”formativ” och ”summativ” infördes har de emellertid blivit ”allmångods” och använts med flera andra betydelser, menar Scriven.

Summativ bedömning

Summativa bedömningar av elever har alltså det gemensamt, att de är kopplade till någon form av betygssättning eller rangordning av elevernas prestationer. Deras grundläggande syfte är därför att generera en så tillförlitlig bild som möjligt av sådana kvalitativa och kvantitativa aspekter av elevernas lärande som är relevanta i förhållande till undervisningsmålen, och att värdera och tillskriva dessa någon form av betyg eller omdöme.

¹⁷ i en av de första skrifterna om utvärdering, den amerikanska ”The Methodology of Evaluation – min översättning

Certifikation

Summativ bedömning av individer kan syfta till certifikation av framgångsrikt fullbordade studier, dvs till att intyga att personen ifråga besitter en viss kompetens. Bedömning som syftar till certifikation fyller funktionen att eliminera okvalificerade sökande till utbildningar och yrken av sådana slag, som förutsätter en viss kompetens eller vissa förkunskaper redan från starten (Messick, 1986).

Differentiering

Summativa bedömningar används också för att differentiera individer, något som är aktuellt i de fall då antalet kvalificerade sökande till en utbildning eller ett yrke överstiger antalet tillgängliga platser. Inom utbildningssystemet är det svenska högskoleprovet och det amerikanska provet SAT¹⁸ är exempel på prov med uteslutande differentierande syfte. Resultaten på dessa prov är inte avsedda att förmedla en bild av individens kompetens i sig, utan enbart att skilja presumtiva sökande åt.

Betyg och examina kan, i sin egenskap av intyg på de enskilda individernas meriter, betraktas som en form av "valuta" som kan användas i konkurrensen om önskade positioner på utbildnings- och arbetsmarknaden.

Kompetens och kvalifikation

Ett annat syfte med summativ bedömning är kopplat till skolans ansvar gentemot elever och föräldrar att eleverna verkligen får den utbildning som de har rätt till – exempelvis har elever i det svenska skolsystemet rätt till så mycket undervisning som fordras för att de ska tillägna sig tillräckliga kunskaper för att få betyget godkänd i de skärnämna (matematik, svenska och engelska). Bakgrunden till detta är den ursprungliga idén bakom grundskolan, vilken också har kommit att utsträckas till gymnasiet, nämligen

¹⁸ SAT (Scolastic Achievement Test) är ett prov som används för urval till gymnasieskolan

gen att den å ena sidan ska utgöra en allmän medborgarutbildning, som gör det möjligt för var och en att ta tillvara sina demokratiska rättigheter och aktivt delta i samhällslivet, å andra sidan lägga en kunskapsmässig grund för yrkesliv och fortsatta studier.

Kvalitetssäkring och samhällsekonomiska intressen

På systemnivå används summativa bedömningar i sådana skeden då man ska besluta om exempelvis en kurs eller ett program ska utvecklas vidare eller läggas ner (Worthen & VanDusen, 1991). Andra exempel på summativa utvärderingar är sådana som syftar till att ge underlag för jämförelse utbildningskvaliteten vid olika skolor, distrikt och nationella utbildningssystem. Så kallade nationella utvärderingar är en typ av utvärdering av den nationella kunskapsnivån och läroplanen. På basis av svaren från ett stort antal elever (i princip hela årskullar) kan man dels sammanställa en totalbild av den nationella kunskapsstandarden i olika ämnen och årskurser, dels jämföra kunskapsstandarden mellan olika grupper.

Bedömningarna syftar då till att dels garantera att eleverna att de får en god och likvärdig utbildning (dvs att kommunerna uppfyller skollagen) och att se till att en nationell läroplan existerar i praktiken (dvs att det finns ett minimum av innehåll, som alla elever får del av oavsett region eller skola), dels till att försäkra staten om avkastning på den ekonomiska investering som man gör i skolan.

Formativ bedömning av enskilda elever

När det gäller enskilda elever, skriver Scriven (1967, i den text där begreppet lanseras), är den formativa bedömningens syfte dels att påverka deras beteende i linje med de uppsatta undervisningsmålen, dels i att korrigera fel. Schweizaren Philippe Perrenoud (1991) menar, att all bedömning som hjälper eleven att lära sig och utvecklas är formativ. Formativ bedömning bygger, säger han, på att ”man inte ger sig” förrän man har lärt sig och behärskar de kunskaper och färdigheter som undervisningen har syftat till.

Fransmannen Bonniol (1991) menar, att målet med den formativa bedömningen bör vara, att eleven efterhand ska utveckla sin egen förmåga att utvärdera sin behållning och utveckla sina lärandestrategier. Det är alltså elevens strategier, som ska korrigeras – ”the procedures, the uses and the way he looks at and tackles his world” (ibid., s. 134). Därför, menar Bonniol, är det processen, som ska stå i fokus vid den formativa bedömningen och inte resultatet : det är betydligt lättare att ”rätta” ett felaktigt svar än att bedöma den process om lett fram till svaret. Ofta säger svaret inte mycket om hur eleven har resonerat och kan då heller inte vara till stor nytta när det gäller att hjälpa eleven vidare -därför får eleven bara höra att han borde ha läst mer, ansträngt sig mera, lyssnat bättre etc . Det är inte heller alltid lätt för eleven att utifrån ett svar dra sig till minnes hur han resonerade vid det tillfälle då han svarade, menar Bonniol.

Bonniol (ibid.) påpekar vidare, att det pedagogiska skälet för att man ”rättar” elevers felaktiga svar oftast är att man vill att eleven ska lära sig något (vad som blev fel, och vilket sätt att tänka och göra som hade varit mera framgångsrikt). Denna erfarenhet kan göra att eleven, nästa gång han eller hon ställs inför en liknande uppgift, handlar annorlunda, d v s justerar sitt eget beteende. Om eleven faktiskt gör det är emellertid en annan fråga, Bonniol poängterar det faktum att eftersom ingen person utifrån kan justera någon annans beteende, så kan läraren egentligen inte ändra sina elevers beteende – det kan de bara göra själva.

Perrenoud (1991) pekar på vikten av, att formativa bedömningar utformas och genomförs utifrån sitt specifika syfte; om de utformas enligt de kriterier som är viktiga vid summativ bedömning förlorar de i effektivitet och kvalitet: För att formativ bedömning ska bli effektiv, menar Perrenoud, ska insamling och bearbetning av olika slags data om elevernas kunskaper inte styras av strävan att vara systematisk eller rättvis, eller att passa vissa utvärderingssystem, utan

(...) förse läraren med den information som han behöver för att kunna påverka villkoren för elevernas lärande eller elevernas inlärningsprocesser på ett effektivt sätt – varken mer eller mindre. (ibid., s. 95)

I litteraturen framstår följande formativa syften som de mest vanliga när det gäller bedömning på individnivå, d v s av enskilda elever:

- att diagnostisera problem och identifiera elevens behov av hjälp och stöd i lärandet
- att inventera elevers förkunskaper/förförståelse för att kunna planera undervisningen
- att hjälpa elever att reflektera över sitt eget lärande och bli varse om sin progression och sina svårigheter och behov
- att vägleda elever i utvecklingen av effektiva inlärnings- och tankestrategier
- att definiera den aktuella disciplinens kunskap för eleverna, d v s markera diskursens gränser och överföra kulturella begrepp och förståelsehorisonter
- att motivera elever i skolarbetet

De engelska forskarna Black & William (1998), drog av den meta-analys av 40 kontrollerade studier av fall där man försökt utveckla provens formativa egenskaper, följande slutsatser:

- elevernas genomsnittliga studieresultat (mätt i prestationer på de vanliga nationella proven) förbättrades påtagligt
- den största förbättringen skedde i de svagast presterande grupperna, d v s skillnaden mellan hög- och lågpresterande elever minskade
- att eleverna själva är involverade i bedömningsprocessen har goda effekter på lärandet
- bedömning och självbedömning påverkar elevernas motivation och självvärdering
- eleverna kan bara bedöma sig själva om de har en klar uppfattning om undervisningsmålen
- om eleverna enbart får poäng eller betyg, så gynnas deras lärande inte

Shepard (2000) beskriver en studie där matematiklärare gav kvalitativ feed-back – kommenterade fel och dåliga strategier, föreslog förbättringar och betonade förståelse snarare än ytlig kunskap. Studien visade att eleverna blev signifikant duktigare i matematik, att deras attityder till ämnet förbättrades och att pojkarnas ursprungliga försprång minskade. Shepard betonar att

eleverna behöver förstå bedömningsprocedurerna och -målen som medel för att utveckla sin förmåga att se sitt eget lärande:

the features of excellent performance should be so transparent that students can learn to evaluate their own work in the same way that their teachers do. (Shepard, *ibid.*, s. 31)

Vidare menar Shepard, att eftersom lärande är en social process, bör bedömningen göras i grupp: då socialiseras eleverna in i ämnets tankemönster och tränar sig i att förklara sina resonemang och få feedback på sin kompetensutveckling av sina kamrater. Dessutom krävs att det traditionella förhållandet mellan lärare och elev öppnas och elevens perspektiv erkänns i klassrummet.

Ingen annan känd strategi för att förbättra utbildningskvalitet har, enligt Black & William (*ibid.*) så stor effekt som utvecklingen av formativ bedömning. Shepard påpekar dock att olika elever har olika lätt att ta till sig de goda effekterna av formativ bedömning. Vissa elever är uppgifts- eller prestationsorienterade och andra inlärningsorienterade och det är de inlärningsoroentade eleverna som lättast accepterar formativ bedömning. Elever som är "fångna i identiteten som dålig elev och bråkstake" och vars mål är att "klara soldagen utan några större katastrofer" (Perrenoud, i Shepard, 2001, s.28), utvecklar strategier för att dölja sina bristande kunskaper. En förutsättning för att formativ bedömning ska lyckas, är därför att läraren motverka sådana dysfunktionella strategier hos sina elever, menar Shepard (*ibid.*).

Formativ bedömning och undervisning

De kontinuerliga utvärderingar eller självvärderingar, som lärare dagligen gör av sin undervisning för att kunna planera kommande undervisningpass och individuella insatser är också ett slags formativ bedömning. Om lärares bedömning av sin egen undervisning sker utifrån tydliga kriterier och resultatet på ett systematiskt sätt används i utvecklingen av den egna undervisningen, kan denna kontinuerliga process av utvärdering – modifiering – utvärdering sägas utgöra en form av aktionsforskning.

Cambourne & Turbill (1990, refererad i Shepard, 2000) menar också att de metoder som föreslås av Lincoln & Guba för att försäkra sig om kvaliteten i data och analys i kvalitativ forskning i princip är tillämpliga också i klassrummet när det gäller bedömningen av elevers kunskaper. Det vanliga torde dock vara att den cykliska utvärderings-handlings-cykeln till stora delar är resultatet av lärarens informella observationer, intuitiva tolkningar och spontana beslut, fattade ”i farten”¹⁹.

Formativ bedömning i mångkulturell utbildning

Utvecklingen av formativa bedömningsmodeller har ibland använts som ett verktyg för att höja utbildningskvaliteten för elever, vars kulturella bakgrund gör att de ofta kommer till korta i utbildningssystemet. Det handlar bl a om elever som tillhör etniska och språkliga minoriteter. I begreppet ”mångkulturell utbildning”, som börjar få stor spridning, vidgas dock resonemanget ofta även till elever ur arbetarklassen och flickor²⁰. I förhållande till den kultur som genomsyrar det västerländska samhället och även dess utbildningssystem och som antas vara den ”manliga, vita, medelklassens kultur”, innebär ju varje annan kulturell tillhörighet en form av avvikelse (se Banks, 1999 och Tesfahuney, 1999).

¹⁹ Jan Bengtsson (1996) menar, att lärare sällan har tid att ”stiga ur” undervisningskontexten för att ”reflektera” över sin undervisning. De tusentals beslut, som den genomsnittlige läraren fattar under en skoldag, vilar därför på lärarens förmåga att göra blixtnabba analyser av sammansatta situationer på grundval av data från en mängd olika källor. Denna till stor del intuitiva förmåga anses vara en bärande dimension i lärarkompetensen.

²⁰ I Sverige kan dock flickor som grupp knappast sägas komma till korta i utbildningssystemet, om man som mått på i vilken utsträckning de drar nytta av och kommer till sin rätt i skolan tar statistik över betyg, specialundervisningsinsatser och övergång till högre studier. Sådan statistik talar genomgående för att flickorna har ett överläge i skolan i förhållande till pojkar (Skolverket, 2001. Det ”överläge”, som avspeglas i denna typ av statistik, innebär dock inte nödvändigtvis att innehållet och arbetsformerna i skolan generellt faktiskt gynnar flickor eller speglar deras intressen och erfarenheter på ett bättre sätt än pojkars. Forskning som inriktats på att undersöka kommunikationsmönster och ideologiska budskap i undervisningen, visar snarare att undervisning och läromedel ofta bygger på en maskulin norm (t ex Gipps, 1994; Weiner & Berge, 2000 och Tesfahuney, 1999).

På vissa skolor utgör regelbundna möten, där lärarna diskuterar sina elevers arbeten och försöker utveckla gemensamma bedömningskriterier, ett viktigt medel för gemensam reflektion om kvalitet, mål och medel i undervisningen (Wolf & Reardon, 1996; BonillaBowman & Gordon, 1996; Darling-Hammond & Aness, 1996). Wolf och Reardon (1996), som har varit engagerade i att utveckla formativa bedömningsmodeller i multikulturella amerikanska skolor, beskriver hur sådana möten skiljer sig från traditionella betygskonferenser:

Målet är inte att komma fram till ett reliabelt resultat. Det är framför allt att utveckla en gemensam uppsättning av dimensioner för att diagnostisera och ge återkoppling på elevernas arbete (ibid., s. 22).

Här handlar det alltså om att konstruera gemensamma kriterier för att bedöma kvaliteten i elevernas arbete – yttre kriterier. En strategi, som beskrivs av Gordon & Bonilla Bowman (1996), för att göra eleverna medvetna om vad som utgör "kvalitet" i olika typer av arbeten, och hur kvalitet kan uppnås, är att involvera eleverna i framtagandet av bedömningskriterier:

I många fall leder den växande förståelsen och införlivandet av kriterier, som blir synliga genom elevernas egna bedömningar av sitt arbete och deras reflektioner, till en personlig arbetsdefinition av kvalitet (ibid., s. 37).

Genom att låta eleverna analysera sina egna, varandras och professionella arbeten, hoppas man att de ska bli medvetna om och utveckla både "inre" och "yttre" kriterier. Idén är att elevernas arbete med kriterier, ska hjälpa dem att finna sätt att hantera och uttrycka kunskap som relaterar både till den egna kulturella, lingvistiska, klass- och könsmässiga tillhörigheten och till majoritetskulturens krav. I längden hoppas man att det här sättet att arbeta med bedömning ska bidra till att elever ur "minoritetsgrupper" får en utbildning som å ena sidan synliggör och utvecklar det specifika i deras kulturella identitet, å andra sidan hjälper dem att bemästra den "dominerande kulturens grammatik" (Darlington-Hammond & Aness, 1996). Wolf och Reardon (1996) talar om detta som en kulturell tvåspråkighet, "biculturalism".

Vad är det som ska bedömas?

Mot vad riktar sig lärares bedömning av elever? Vad en lärare väljer att fokusera vid ett specifikt provtillfälle beror naturligtvis på syftet med bedömningen och på undervisningens specifika syfte. I detta kapitel reds olika begrepp ut, som gäller elevers lärande och resultatet av detta, och jämförs med varandra. Kunskaper, förståelse, prestationer och kompetens är några av de begrepp som behandlas.

Kunskap

Vad det innebär att ha kunskap, hur giltig kunskap om världen kan nås och hur man kan vet att man vet något, är filosofiska frågor, som besvaras på olika sätt av t ex empirister och rationalister. Hos exempelvis Platon, Aristoteles och Kant är kunskap liktydigt med "sann, välgrundad iakttagelse" (Lübcke, 1988). Många senare filosofer har visat att denna definition av olika skäl är problematisk. Jag ska dock inte här fördjupa mig i filosofiska resonemang om kunskap. Istället ska jag utifrån delar av huvudbetänkandet från 1994 års läroplanskommitté (SOU 1992:94) och olika så kallade taxonomier visa hur man i några pedagogiska sammanhang har definierat "kunskap".

Kunskapsbegreppet i LPO 94

Ingrid Carlgren skriver i ett särtryck (Bildning och kunskap) till detta betänkande, att den

utvecklas i ett växelspel mellan vad man vill uppnå, den kunskap man redan har, problem man upplever med utgångspunkt i denna, samt de erfarenheter man gör. Kunskapen fyller en funktion, löser ett problem eller underlättar en verksamhet. (Carlgren, 1994, s. 29)

Vidare skriver hon att kunskap inte bara är kognitiv, utan också "sinnlig och praktisk". Varje individs kunskap är dessutom bara till en del synlig och formulerad. En stor del av individens kun-

skaper är ”tysta och finns som en underförstådd bakgrundskunskap” (ibid., s. 31). Den ”tysta kunskapen” är, menar Carlgren, oundgänglig exempelvis när man ska lösa problem eller göra bedömningar.

I betänkandet, såväl som i den därpå följande läroplanstexten, delas kunskapsbegreppet in i fyra delaspekter, nämligen:

- *Faktakunskaper*, vilka av Carlgren beskrivs som ”kunskap som information, regler och konventioner” (ibid., s. 32). Faktakunskaper har en kvantitativ dimension genom att den kan mätas i ”mer eller mindre”. Det finns ingen åtskillnad mellan ytlig och djup faktakunskap.
- *Förståelsekunskap*. ”Att förstå är att begripa, att uppfatta meningen eller innebörden i ett fenomen”, skriver Carlgren (ibid.). Språket är fundamentalt för förståelsen – ju mera utvecklat språket är, desto mer nyanserade uppfattningar och komplexa tankekonstruktioner är möjliga. Fakta och förståelse är beroende av varandra. ”Förståelse” är en teoretisk kunskapsform.
- *Färdigheter*. Dessa utgör, till skillnad från förståelse, en praktisk kunskapsform, menar Carlgren. Att ha färdigheter innebär att veta hur något ska göras och att kunna göra det.
- *Förtrogenhetskunskap* är ungefär detsamma som ”tyst kunskap”. Carlgren menar att den ofta är sinnlig. ”Vi ser, luktar, känner och ’vet’ när något skall avbrytas eller påbörjas” (ibid., s. 33). Att känna sig hemma i och behärska t ex olika sociala eller språkliga situationer förutsätter förtrogenhetskunskap. Förtrogenhetskunskap tillägnar man sig i stor utsträckning genom socialisation, d v s genom att delta i olika sociala sammanhang och införliva de värderingar och de ”förståelsehorisonter” som de andra i gruppen delar (jämför t ex Laves och Wengers bok *Legitimate Peripheral Participation*). Carlgren (ibid., s. 33), skriver: ”Genom att delta i praktiska verksamheter lär vi oss reglerna i dessa verksamheter.”

I betänkandet (och, senare, i läroplanen) betonas att indelningen i de fyra kunskapsformerna ovan inte ska tolkas som att de i verkligheten uppträder åtskilda, eller i "ren" form. I en persons kunnande utgör de integrerade aspekter, som framträder olika starkt beroende på person och situation. De fyra kunskapsformerna ska heller inte uppfattas som hierarkiska, d v s att den ena skulle vara en mer avancerad form än den andra.

Taxonomier

Den tidigaste och mest kända taxonomin är Benjamin Blooms från 1956. Denna var ursprungligen avsedd att hjälpa lärare att göra sina prov mera ändamålsenliga genom att lära sig urskilja olika slags kunskaper (Phye, 1997). Blooms taxonomi är, till skillnad från läroplanens indelning ovan, hierarkisk och indelad i olika nivåer. Den typ av kunskap, som Bloom ser som den enklaste och som han därför hänför till den lägsta nivån, betecknar han just "kunskap". Denna kunskap saknar struktur för individen och kan jämföras med fakta utan inbördes samband eller mening. Därefter följer i tur och ordning nivåerna "förståelse", "analys", "tillämpning", "syntes" och "värdering", där den sista, "värdering", alltså ses som den högsta kognitiva nivån (ibid.).

Blooms taxonomi har dock utsatts för en del kritik av moderna forskare (ibid.). En anledning är att de olika nivåerna med nu aktuella sätt att se på kunskap och lärande inte uppfattas som distinkta – jämför t ex läroplanens kunskapsbegrepp ovan. Exempelvis ser man numera tillämpning som den säkraste indikationen på förståelse, varför dessa två "nivåer" flyter samman till en.

Man har också funnit det diskutabelt att oavsett hur eleverna tolkar frågan, förutsätta att de tankeverksamheter som benämns "värdering" och "syntes" generellt kan sägas representera mera avancerade kognitiva processer än det som Bloom benämner som kunskap respektive förståelse (Stiggins i Phye, 1997). Stiggins illustrerar detta med exemplet, att om man frågar ett barn vilken av två böcker han eller hon tycker är bäst, så finns möjligheten, att barnet ger ett svar som baserar sig på analytiskt

tänkande och syntes. I sådana fall kan svaret visa på barnets "higher-order thinking skills". Ett allvarligt problem med denna fråga i egenskap av instrument för att mäta sådana kognitiva färdigheter är emellertid, att den faktiskt inte fordrar ett sådant tänkande av barnet; barnet kan precis lika gärna svara att den ena boken är bäst av den anledningen att det var den som han eller hon tyckte mest om. Stiggens menar, att det är sättet på vilket frågan ställs och de krav, som frågans formulering ställer på svaret, som avgör hur komplext tänkande, som frågan kan mäta. Att en och samma individ kan lösa en uppgift på ett kognitivt mer avancerat sätt vid ett tillfälle än vid ett annat, beroende på hur hon har uppfattat uppgiften och meningen med den och beroende på hur hon (eventuellt som en konsekvens av detta) har gripit sig an den, visas bl a i en studie av Säljö (1975).

Bloom problematiserar inte heller relationen mellan elevers tänkande och deras prestation i provsituationen – bedömningen av elevens kunskaper sker uteslutande på basis av prestationen. Enligt Phye (1997), är det ointresse, som Blooms taxonomi speglar, för "vad som sker inuti elevernas huvud", typiskt för 1950-talets behaviouristiska inlärningssyn, i vilken lärande betraktas som ett beteende snarare än en inre process.

En modell, som fokuserar de inre, kognitiva processerna och som ofta används idag, är Quellmalz "framework for reasoning skills" (Quellmalz & Hoskyn, 1997). Denna syftar till bedömning av elevers förmåga att resonera, d v s tänka och argumentera logiskt och självständigt. Quellmalz & Hoskyns indelning i fem komponenter 1) återgivning 2) analys 3) jämförelse 4) slutledning och 5) värdering är inte, som Blooms, hierarkisk. Det innebär, att "återgivning" inte automatiskt definieras den minst sofistikerade strategin, och att "värdering" inte självklart ses som den mest sofistikerade. Vilken strategi som är mest avancerad beror på situationen och på frågan eller uppgiften. Quellmalz och Hoskyn (ibid.) framhåller att tänkandet är beroende av innehållet och sammanhanget, det är alltid ett tänkande om något specifikt, i en specifik situation.

Phye (1997, s. 48 ff) använder en tredje modell för att beskriva kunskapsbegreppet. Han delar i det i tre former, som han benämner 1) deklarativ kunskap, 2) processkunskap och 3) strategisk kunskap. Den deklarativa kunskapen handlar om att ”veta vad”. Exempel på deklarativa kunskaper är ordkunskap, fakta, begrepp och andra ”informationsbitar” av den art som eleverna har lagrade i sitt långtidsminne. Processkunskap handlar om ”att veta hur”, och att ha processkunskap innebär att kunna gruppera, kombinera, rekonstruera och assimilera deklarativ kunskap, så att den kan användas i tillämpning. I akademiskt lärande, så innebär detta att kunna transformera information till planer, strategier, idéer och liknande (ibid.).

Om den deklarativa kunskapen handlar om att ”veta vad” och processkunskapen om att ”veta hur”, så handlar den *strategiska kunskapen* om att ”veta när” olika sorters tankestrategier ska tillämpas. ”Strategisk kunskap” betecknar förmågan att veta när och hur den deklarativa kunskapen och processkunskapen ska användas för att åstadkomma ”inlärningsresultat”, menar Phye (ibid.).

En fjärde taxonomi som har fått stor internationell spridning är den så kallade SOLO-taxonomin (Biggs & Collis, 1982). SOLO-taxonomin är dock enbart avsedd som ett instrument för att bedöma elevers provprestationer, och kommer därför att beskrivas i avsnittet ”Att värdera elevers prestationer” längre fram.

Nedan har de tre övriga taxonomierna, jämte den indelning av kunskapsbegreppet som presenteras i LPO 94, placerats i en tabell för att de olika kunskapsbegreppen ska bli mer åskådliga och lättare att jämföra.

BLOOM	QUELLMALZ	PHYE	LPO 94
Kunskap	Återgivning	Deklarativ kunskap	Fakta
Förståelse	Analys	Processkunskap	Förståelse
Syntes	Jämförelse		Färdigheter
Tillämpning	Slutledning		
Jämförelse	Värdering		
Värdering		Strategisk kunskap	Förtrogenhet

De olika taxonomierna och läroplanen konstruerar, som vi har sett, alltså kunskapsbegreppet något olika med avseende på i vilken mån olika slags eller former av kunskaper kan ses som åtskilda i verkligheten och hierarkiskt ordnade. Blooms taxonomi skiljer sig därvid mest från de övriga genom att hävda en tydlig åtskillnad och en tydlig hierarki mellan de olika nivåerna.

Medan läroplanens kunskapsbegrepp, genom att ta upp ”färdigheter” som en kunskapsform innefattar även praktiska kunskaper, avser de tre taxonomierna enbart kunskap som kognitivt fenomen. I läroplanen beskrivs ”färdigheter” nämligen som en praktisk kunskapsform, till skillnad från ”förståelse”, vilken beskrivs som en teoretisk kunskapsform. I Nationalencyklopedin (1995), liksom Norstedts svenska ordbok (1999) definieras ”färdighet”, *som förmågan att utföra något i praktiken (ibland med hjälp av teoretiska kunskaper)*. I senare års inlärningsforskning är emellertid sammansättningen ”kognitiva färdigheter” vanlig. Quellmalz & Hoskyns (1997) begrepp ”reasoning skills” handlar också om hanterandet av teoretiska uppgifter snarare än praktiska. Färdighetsbegreppet har alltså vidgats och gränsen mellan praktiska och teoretiska kunskaper delvis lösts upp.

Enligt Quellmalz & Hoskyn (ibid.) delas pedagogernas intresse för ”kognitiva färdigheter” även av filosofer och psykologer. ”Kritiskt tänkande”, ”problemlösning”, ”strategiskt resonerande” och ”reflekterande tänkande” är begrepp som, enligt Quellmalz & Hoskyn, används inom olika discipliner, men refererar till väsentligen samma sak.

Kompetens

”Kompetens” betyder enligt Nationalencyklopedin (1995) och Norstedts svenska ordbok (*tillräckligt*) *god förmåga (för viss verksamhet)* och är synonymt med duglighet eller skicklighet.

”Kompetens” kan också, enligt samma källor, specifikt syfta på *egenskapen att uppfylla de formella kraven för att i ex inneha en viss tjänst: gymnasiekompetens*. Begreppet ”kompetens” används alltså för att beteckna specifika kombinationer av olika slags ”kunskaper” och ”färdigheter”, exempelvis ”teoretisk” och ”praktisk”

kunskap, eller av "tyst" och manifest kunskap. Begreppet ställs då i relation till "kunskap" i en mera avgränsad mening, exempelvis "faktakunskaper" (jämför Bloom, 1956). "Kompetens" kan emellertid också kontrasteras mot "behållning", d v s kunskap, som tillägnats i något visst sammanhang. Enligt Messick (1984) refererar "kompetens" till vad en person vet och kan göra inom ett ämnesområde oavsett hur eller när dessa kunskaper eller färdigheter har förvärvats.

När syftet med bedömning är certifikation eller ackreditering (betygssättning), d v s att ge den studerande någon form av bevis för att han eller hon besitter vissa kunskaper och därmed anses kvalificerad för exempelvis ett visst yrke eller fortsatta studier, så är det alltså vederbörandes kompetens inom de aktuella domänerna, som är intressanta. I vilken utsträckning kunskaperna har förvärvats inom ramen för exempelvis den aktuella kursen eller utbildningen är inte relevant.

Behållning och prestation

Med "behållning" avses det som eleven har lärt sig som en följd av planerade undervisningshandlingar (Messick, 1993). För att kunna bedöma behållningen måste elevernas kunskaper i den aktuella domänen inventeras både före och efter undervisningen – något som alltså inte är nödvändigt vid bedömningen av kompetens.

I praktiken är det dock omöjligt (och ofta heller inte relevant) att avgöra i vilken utsträckning som skillnaden mellan den kunskap, som eleven uppvisar före respektive i slutet av en undervisningssekvens, faktiskt är resultatet av själva undervisningen, eller av lärande som har skett oberoende av men tidsmässigt parallellt med undervisningen.

I bedömning som syftar till utvärdering av undervisningen eller till kartläggning av elevernas kunskapsprogression är det elevernas behållning som fokuseras. Att inventera förkunskaper, eller förförståelse är relevant 1) för att öka elevernas medvetenhet om och utveckling av sin förståelse och av sina kognitiva strate-

gier 2) vid bedömningen av exempelvis grupparbete eller problembaserat lärande, där elevernas förmåga att arbeta på ett visst sätt är minst lika viktigt som resultatet av arbetet 3) effektivisera och individualisera undervisningen

Förmåga

Gränsen för vad en person kan prestera antas sättas av hennes förmåga. Begreppet "kognitiv förmåga" används ofta med samma betydelse som "begåvning" eller "intelligens" och utvärderas eller bedöms inte av lärare i klassrummet som en del i det pedagogiska arbetet, utan är främst en angelägenhet för psykologer och forskare (Choppin, 1990). En orsak till detta är att den kognitiva förmågan traditionellt har betraktas som inneboende och konstant. Eftersom man av denna anledning har antagit att intelligensen inte i nämnvärd utsträckning påverkas av pedagogiska insatser, har den inte i första hand setts som en angelägenhet för läraren. Med nyare synsätt på intelligens, där metakognition och utveckling betonas, blir emellertid elevernas "förmåga" inte enbart en förutsättning för den pedagogiska verksamheten, utan kan också ses som föremål för den.

Olika provformer

I detta kapitel behandlas de metoder och instrument, som används för att utvärdera kunskaper, d v s ”själva *proven*”. Anledningen till att jag väljer ordet ”utvärdera” här, trots att det som fokuseras är bedömning på individnivå, är, att jag för tydlighetens skull vill skilja dessa instrument, ”proven som sådana”, från den process, där elevernas ”svar”, eller andra prestationer som är avsedda att utgöra bevis på deras kunskap, *värderas* i förhållande till varandra eller till något kriterium. Värderingsprocessen behandlas i nästa kapitel. Det ska också påpekas, att ordet ”prov” i detta kapitel används med en bredare betydelse än vad som vanligen läggs i det. ”Prov” refererar här till alla sådana uppgifter, skriftliga såväl som praktiska, som ges till elever med avsikten att skapa information om deras kunskaper. Informationen kan vara avsedd för olika mottagare – förutom för eleven själv, exempelvis lärare och föräldrar. Även forskare, statliga utredare och politiker kan, som nämnts, ha intresse av att mäta elevers kunskaper, men då för att dra slutsatser på en principiell eller systemnivå. I detta avsnitt behandlas endast sådana prov som görs i utbildningssammanhang och har betydelse för de lärande själva.

”Konventionella” och ”alternativa” prov

En princip för att kategorisera olika slags prov är med utgångspunkt från vilket slags prestation som de kräver av eleverna. Ska eleverna svara skriftligt? Ska de själva konstruera sina svar eller finns givna svarsalternativ? Ska de lösa ett problem?

I dagligt tal skiljer man ofta mellan ”konventionella” och ”alternativa” prov. Vad är då utmärkande för konventionella respektive alternativa prov? Själva beteckningen konventionell syftar ju på att det handlar om den vanligaste sortens prov, prov som följer traditionen, som betraktas som normal praxis. Begreppet ”alternativ” måste då i sammanhanget beteckna de sorters prov

som inte kan räknas till de vanliga, traditionella, prov som avviker från vad som betraktas som normal praxis. Men vad som är vanliga och etablerade metoder för att pröva elevernas kunskaper varierar ju mellan olika ämnen. Exempelvis idrott, historia, kemi och matematik har alla sina etablerade metoder. Vad som är en konventionell metod för kunskapsbedömning i ett ämne, kan därför betecknas som ”alternativ” i ett annat. Eftersom begreppen konventionella och alternativa prov ändå är så vanligt förekommande (också i forskningslitteraturen) och det faktiskt tycks finnas något samförstånd om betydelsen, väljer jag trots det att ta upp dem här.

Själva arketyper för ”prov” i vårt samhälle, och i de teoretiska skolämnena troligen också den vanligaste formen, är ett papper med ett antal skriftliga frågor, som ska besvaras av eleverna individuellt, på en viss tid och under övervakning. Proven kan ha fasta svarsalternativ, och kallas då ibland ”objektiva”, eller bygga på att eleverna själva formulerar sina svar. I klichén ingår också att frågorna är av reproducerande karaktär, vilket innebär att de kan besvaras antingen rätt eller felaktigt. Med ”konventionella prov” avses ofta prov som uppfyller de flesta av dessa kriterier.

Som ”alternativa” betecknas då istället ofta sådana prov som

- snarare än av skriftliga frågor, består av praktiska uppgifter, där eleven får demonstrera sin kunskap i handling, eller genom någon produkt som är resultatet av deras handlingar
- ofta är inriktade mot divergent tänkande, vilket innebär att svarsutrymmet inte definieras i förväg, utan sätten att lösa uppgifterna med tillfredsställande resultat i princip kan variera i det oändliga;
- är inriktade mot högre kognitionsnivåer och metakognition – processkunskap, strategisk kunskap snarare än faktakunskap
- inkluderar även kreativa och affektiva aspekter på kunskap

En av de vanligaste formerna av ”alternativa prov” är sk tillämpningsprov (Cunningham, 1998). Principen för dessa är att eleverna bedöms på basis av hur de praktiskt handlar i en situation, vars bemästrande fordrar att eleven använder och

därigenom visar upp den aktuella kompetensen. Det beteende, eller den kunskap, som man är intresserad av, observeras och bedöms alltså direkt. Tillämpningsproven skiljer sig från sådana prov som består av ett urval av frågor, dvs som är avsedda att utgöra ett stickprov på den kunskap man vill mäta och där man istället för att observera elevernas kunskap direkt, drar slutsatser om vad de kan på basis av hur de har svarat på dessa ”stickprovsfrågor”. Uppgifterna i tillämpningsprov måste sammanfalla med de mål som man har haft med undervisningen. När man vill använda denna provform är det därför särskilt viktigt att syftet med undervisningen är väl definierat: ”Man måste fråga sig själv vad man förväntar sig att eleverna ska kunna göra vid kursens slut och det är detta som ska vara provets fokus.” (Cunningham, *ibid*, s. 132). Tillämpningsprov passar bäst, menar Cunningham, när målet med undervisningen är att eleverna ska kunna utföra något.

Cunningham använder exemplet att om en lektion, vars syfte är att eleverna ska lära sig skriva noveller, bedöms med hjälp av ett tillämpningsprov, så innebär det att eleverna faktiskt får skriva en novell och att det är på basis av detta arbete som man drar sina slutsatser om elevernas kunskaper i novellskrivning. Ett indirekt sätt att testa samma sak hade kunnat utgöras av ett prov där eleverna ska besvara frågor om novellskrivandets tekniker och kvalitetskriterier. Ett annat exempel på tillämpningsprov är den uppgift, som gavs till pedagogikstuderande vid Göteborgs universitet i början av 1970-talet för att testa deras kunskaper om fenomenet ”betingning”; istället för att ge studenterna en skriftlig tentamen med frågor om detta fenomen, så gav man dem varsin så kallad Skinnerbox (en låda där råttan kunde få mat genom att utföra vissa beteenden) och varsin rätta, som studenterna fick i uppgift att genom ”betingning” lära att bete sig på vissa sätt. Detta exempel visar för övrigt, att undervisning i pedagogikämnet inte alltid är en gestaltning av de teorier som den handlar om. Den direktet i bedömningsprocessen, som innebär att inga slutsatser ska behöva dras i tolkningen av resultatet av ett tillämpningsprov, är tillämpningsprovets kanske mest centrala egenskap.

I engelsk litteratur används ofta termen ”tillämpningsprov” (eng. *performance assessment*) synonymt med ”alternativa prov” och ställs då emot ”konventionella prov”, något som dock riskerar att leda till ytliga slutsatser. I följande citat av Madaus & O’Dwyer sätts vad som idag ofta betecknas som alternativa provformer i ett historiskt perspektiv:

Alltsedan kunskapsmätning började användas systematiskt i Kina år 210 f Kr, har det endast funnits fyra sätt att testa kunskaper inom ett område. För det första kan man be personen ge muntliga eller skriftliga svar på ett antal frågor (t ex essäfrågor, frågor som besvara med ett ord eller en fas, eller muntliga förhör). För det andra kan man be personen att framställa en produkt (t ex en mapp med arbetsprover, en vetenskaplig uppsats, en stol, en stycke slipat glas). För det tredje kan man be personen att utföra en handling, som ska utvärderas i förhållande till särskilda kriterier (t ex genomföra ett kemiskt experiment, läsa högt från en bok, reparera en för-gasare, köra en bil). Slutligen kan man låta den som ska bedömas välja ett svar på en fråga eller ett problem bland flera alternativ (t ex flervals- eller ”sant-/falskt-uppgifter”). Vi använde en vid definition av ”tillämpningsprov”. Med ”tillämpningsprov” avser vi prov som fordrar att den vars kunskaper ska prövas antingen formulerar egna svar, eller genomför eller framställer något som kan bedömas. Således omfattar ”tillämpningsprov” de första tre sätten och utesluter den fjärde. (Madaus & O’Dwyer, 1999, s. 689)

Alla prov vars besvarande kräver någon typ av skapande insats av eleverna kan alltså betraktas som tillämpningsprov i någon mening. Endast prov med objektiva svarsformat faller utanför. Så dominerande är denna typ av prov knappast, åtminstone inte på klassrumsnivå och åtminstone inte i Sverige, att det kan anses motiverat att definiera alla andra slags prov som ”alternativa”.

Objektiva prov

Exempel på objektiva svarsformat är (1) flervalsfrågor (2) påståenden som ska markeras med endera ”sant” eller ”falskt” och (3) uppgifter där olika alternativ ska paras ihop. Prov med objektiva svarsformat är lämpliga i fall då undervisningen har syftat till att eleverna ska tillägna sig vissa termer och teorier menar Cunningham, (1998). Anledningen är att de kunskaper man då vill bedöma härrör från konvergent, och inte divergent tänkande, vilket är en förutsättning för att objektiva prov ska kunna

användas (ibid.). Däremot utesluter inte objektiva provformat mätning av kunskaper på högre kognitionsnivåer, menar Cunningham. Så kan flervalssprov med fördel användas för att bedöma en persons förmåga att exempelvis resonera logiskt, något som det svenska högskoleprovet innehåller flera exempel på.

Två fördelar med att använda ”objektiva prov” är att de i förhållande till andra slags prov ger resultat som är lätta att jämföra och att bedömningen tar lite tid i anspråk. En invändning mot det objektiva provformatet är dock, att det inte ger någon information om *vad* eleven faktiskt kan eller vet, och vad hon *inte* kan eller vet, utan bara vilka frågor hon har svart rätt och fel på. Så länge provet gäller kunskaper på låga taxonomiska nivåer (där det i princip handlar om att känna till vissa fakta eller termer) är detta inget problem, men vid utvärdering av kunskaper på högre taxonomiska nivåer och mera sammansatt kunskap, så är det svårare att veta vad det egentligen är som svaren visar (Cunningham, ibid.).

Långa och korta svar

Prov som eleverna själva ska besvara skriftligt, kan fordra långa utläggningar eller enstaka ord. Man brukar skilja mellan prov som består av ”kortsvarsfrågor” och ”essäfrågor”. De senare fordrar längre svar, allt från någon enstaka mening till flera sidor (ibid.), alternativt numeriska beräkningar, algoritmer och formler. ”Kortsvarsfrågor” besvaras med ett ord eller en mening, eller består av ofullständiga meningar, som ska kompletteras (eller av numeriska resultat, formler etc.). Gemensamt för prov som består av ”kortsvarsfrågor” och s k objektiva prov är, att de endast har ett godtagbart svar och därför endast kan mäta konvergent tänkande.

När det gäller så kallade essäfrågor, så skiljer Cunningham (ibid.) mellan tre typer: (1) reproducerande frågor, som är inriktade mot återgivning av enkla eller sammansatta fakta (2) tillämpnings- och problemlösningsuppgifter och (3) kreativa

uppgifter som testar divergent tänkande²¹. Exempel på den andra kategorins frågor är sådana som uppmanar eleverna att pröva en given hypotes, eller dra slutsatser om ett visst fall genom att tillämpa vissa lagar eller regler. Här finns alltså konstruktivt moment med och olika lösningar eller svar kan vara godtagbara, men det finns samtidigt kriterier för vilka sätt som är rätt och vilka som är fel. Den tredje kategorins uppgifter är av sådant slag, att det inte är relevant att bedöma svaren i termer av ”rätt” eller ”fel”. Istället fokuseras frågan om vilka kvaliteter hos eleven, vilken kompetens, som elevens sätt att genomföra uppgiften, indikerar. Exempel på uppgifter av detta slag är prosa- och poesiskrivning. Essäfrågorna kan alltså vara utformade för att mäta såväl konvergent som divergent tänkande.

Autentiska prov

”Autentiska prov” kan räknas som en variant av tillämpningsprov. De vilar nämligen på samma princip om direkthet i bedömningen: inga slutsatser om kunskaper som inte faktiskt har testas accepteras. Därför har de liksom tillämpningsproven, också alltid en konkret dimension. Vad som är särskiljande för de autentiska proven är emellertid, att de har en relevans utanför bedömnings- och utbildningskontexten. Det innebär att genomförandet av provuppgifterna ska vara värdefullt ur något mera hänseende än att visa att eleven kan något. Det ska ha ett värde i sig, t ex estetiskt, praktiskt eller personligt (Newman, 1997). Om det exempelvis gäller en skrivuppgift, så ska den för att betecknas som en autentisk provuppgift dels ha en publik utanför klassrummet, dels ett syfte utöver bedömningen (ibid.).

Den autentiska bedömningspraktiken kan, enligt Newman (ibid.), inte isoleras från den undervisning och det lärande, som man vill utvärdera. Den autentiska bedömningen är nämligen med nödvändighet kopplad till ”autentiskt lärande”, som förutsätter att eleven själv konstruerar kunskap och mening, och tol-

²¹ Jämför Eisners (1985) allmänna indelning av provuppgifter i återgivande, probleminriktade och expressiva uppgifter.

kar, värderar och organiserar information (ibid.). Cunningham (1998) nämner, som exempel på en autentisk provuppgift, att eleverna får i uppdrag att skriva ett brev med syftet att övertyga ansvariga politiker om att ett naturområde ska skyddas. I en sådan uppgift tränas elevernas förmåga att argumentera, samtidigt som denna förmåga kan observeras och bedömas. Undervisning och bedömning flyter alltså ihop.

En typ av bedömningsituation, som ligger nära den autentiska och som förmodligen är vanligare, innebär att eleverna får uppgifter, som *simulerar verkligheten*, dvs eleverna uppmanas att leva sig in i en situation, där vissa (realistiska) förutsättningar är givna och lösa ett problem, som de i princip skulle kunna ställa inför i verkliga livet. Sådana uppgifter kan t ex vara utformade som rollspel eller som planerings- eller konstruktionsuppgifter.

Portfolio

”Portfolio”, eller ”portfölj”, är en modell för kunskapsbedömning, som har fått stor uppmärksamhet på senare år, och som precis som tillämpningsprov och autentiska prov står i nära samband med undervisning och lärande. Vad innebär då ”portfolio” som bedömningsmetod? Enligt Arters o Spandels definition, (refererad i Cizek, 1997, s. 142) består en portfolio av ”en avsiktlig samling av elevarbeten, som visar på elevens ansträngningar, framsteg eller kompetens inom ett eller flera områden. Dessa arbeten kan utgöras av exempelvis textstycken, föredrag inspelade på ljudband, konstnärliga produkter, labrapporter och till och med arbetsblad i matematik”.

Enligt Cunningham (1998) bör en portfolio innehålla ett urval av de arbeten som eleven är mest nöjd med och en motivering till varför dessa produkter har bedömts vara av hög kvalitet samt förslag på hur produkterna skulle kunna förbättras ytterligare. Den bör dessutom innehålla ofullbordade och mindre lyckade arbeten samt arbeten från olika tidpunkter, så att elevens progression blir synlig. Elevernas egna reflexioner över sitt arbete och sin progression, samt deras upplevelse av undervisningen

och av sammanställningen av portföljmaterialet bör också ingå. Det är viktigt att valet av material som ska ingå är elevens eget och att urvalet har skett utifrån en bestämd plan, som har upprättats innan arbetet startar och att materialet är väl strukturerat (ibid.). Genom att återkommande tvingas fatta beslut om vilka arbeten, som ska väljas ut för att ingå i portfolion och samtidigt utveckla och reflektera över de kriterier, som arbetena väljs utifrån, blir eleverna medvetna om sitt eget lärande och sin progression, menar Cunningham (ibid.). Portfolios anses vara pedagogiskt mycket användbara, särskilt för att utveckla elevernas metakognition och inlärningsstrategier, men de kan vara vanskliga när det gäller betygssättning. Enligt Cizek (1997) finns forskning som visar att de har så allvarliga mättekniska brister, att det i deras nuvarande form, inte är tillrådligt att använda dem för några former av bedömningar som medför viktiga konsekvenser för exempelvis elevernas livschanser. Cunningham (1998) menar, att om portfölj ska användas i sammanhang där eleverna rättssäkerhet är viktig, så måste de standardiseras, så att de i någon mån blir jämförbara. Detta skulle emellertid leda till att stora delar av de pedagogiska fördelarna skulle kompromissas bort – exempelvis skulle elevernas känsla av delaktighet och inflytande sannolikt minska.

När passar vilket prov?

Det finns ingen provtyp, som i sig kan betraktas som generellt bättre än andra. För utvärdering av elevernas behållning av undervisningen under en kurs, ska provet väljas så, att det verkligen ”mäter” just de kunskaper, eller just den förmåga, som undervisningen syftar till att utveckla hos eleverna. Men man måste också beakta vad resultatet av provet ska användas till. Om det är ett urvalsprov ställer kraven på rättvisa och rättssäkerhet stora anspråk på provets mättekniska kvaliteter. Om provet har huvudsakligen formativa syften behövs en metod som synliggör inte bara vad eleven kan, utan också vad hon inte kan och vilket slags stöd hon behöver. Man måste också göra ekonomiska och tidsmässiga överväganden inför val av bedömningsform. En form av prov, som uppfyller alla andra kriterier kan

vara så resurskrävande, att läraren inte kan hantera den, eller att det blir ”för lite kvar” till andra undervisningsaktiviteter. När det gäller prov som ges till stora grupper, t ex i omfattande utvärderingar, behöver man oftast av ekonomiska skäl minimera tiden för bedömning av varje prov.

När det gäller prov som är avsedda att stödja lärande och undervisning i klassrummet kan dock bedömning och undervisning ibland utgöra två aspekter av samma aktivitet. Shepard (2000) nämner som exempel ”running records” (ung. ”löpande protokoll”), som en bedömningsform som går att integrera i den reguljära undervisningen. Ett svenskt exempel är Läs- och skrivutvecklingsschemat (LUS) som har börjat användas i svenska klassrum för att dokumentera yngre skolbarns läs- och skrivutveckling. Läraren ”prickar av” i barnets schema när barnet i något sammanhang har visat att hon har erövrat en nytt steg i utvecklingen (t ex spontant läser på förpackningar och skyltar) och använder schemat för att planlägga stimulerande inlärningsaktiviteter för barnet på ”rätt” nivå.

Shepard (ibid.) menar att barn i de lägre åldrarna nästan uteslutande bör testas med informella bedömningsmetoder – som också kallas observationsbaserade – (där t ex föräldrar och lärare diskuterar barnets läs- och skrivutveckling utifrån exempel och situationer), medan barn i mellanstadieåldrarna bör lära sig att delta i mera formella, men autentiska, bedömningsituationer (t ex att presentera projekt för klassen).

Det är viktigt, skriver Shepard (ibid.), att lärare använder varierade bedömningsmetoder, som passar för att utvärdera och främja lärande av olika slag inom olika områden. Poängen med portfolios är att de ger träning i att reflektera och identifiera kvalitet. För att testa vissa färdigheter eller baskunskaper kan det emellertid också vara motiverat att använda prov med fast eller mycket begränsat svarsutrymme. Även formella prov, där alla samtidigt får samma uppgifter att lösa under en viss tid, har sitt berättigande, menar Shepard, och är användbara för att utvärdera kunskaper som eleverna kan behöva visa ”på beställning” i olika situationer – för en blivande lärare kan det vara att

förklara för en förälder hur hennes barn klarar skolarbetet eller fatta ett snabbt beslut om att lägga om undervisningen under en lektion. Det viktiga är att den kombination av olika slags prov som används speglar undervisningens syfte och att eleverna ser denna koppling (ibid.).

Att använda en mångfald av bedömningsformer är också viktigt av det skälet att olika prov mäter olika saker – t ex kan det vara så om man vill bedöma läsförståelse, att återberättande kommer åt en helt annan kompetens än innehållsfrågor. Därför är det både för relevansen i bedömningen och för att stödja utvecklingen av flexibel och robust förståelse, bra att använda mer än en bedömningsform för att utvärdera ett visst undervisningsmål. Elever bör i klassrummet, åtminstone som ett av flera sätt, få bli testade på det sätt som gör att de mest kommer till sin rätt, anser Shepard (ibid.). Men de bör inte alltid enbart testas på det sätt som de själva är mest bekväma med, eftersom god undervisning ska fokusera såväl svaga som starka sidor och sträva mot att eleverna ska utveckla robusta kunskaper, som är användbara i flera olika sammanhang, menar hon. Om olika prov pekar på att eleven brister i någon förmåga bör man sätta in särskilt stöd för att utveckla denna förmåga. Både undervisning och prov ska syfta till att hjälpa eleverna att se kopplingar mellan liknande problem, mellan konkret och abstrakt etc (ibid.). Uppfattningen av kunskap som generaliserbar mellan olika sammanhang kan emellertid ur ett mera långtgående sociokulturellt perspektiv ses som problematisk, eftersom kunskaper ses som integrerade i specifika sociala kontexter och lärande handlar om att socialiseras in i bestämda diskurser, snarare än att utföra en kognitiv akt (Black, 1999).

Att värdera elevers prestationer

I detta kapitel fokuseras bedömningsproceduren, dvs den process där elevens svar eller andra former av prestationer värderas i förhållande till andra elevers prestationer²² eller i förhållande till kriterier eller mål.

Begreppen kriterie- och normrelaterad bedömning tas här upp, liksom holistisk, global, analytisk, atomistisk och dimensionell bedömning. Bedömningsproceduren diskuteras vidare i förhållande till provets form (exempelvis flervalssprov, essäprov, tillämpningsprov) och till de aspekter av lärandet som står i fokus för bedömningen (exempelvis fakta kontra kognitiva strategier). Tekniska modeller för att bedöma elevers kunskap jämförs med konstnärliga och intuitiva. Sist tas några olika sätt för att rapportera eller registrera elevernas resultat upp.

Kriterierelaterad bedömning

När elevers prestationer ska bedömas, så kan det ske i förhållande till antingen normer eller kriterier. ”Kriterier” kan beskrivas som de krav som ställs på en prestation för att den exempelvis ska bedömas som godkänd. Begreppet ”kriterierelaterad bedömning” är emellertid mycket vitt; Hambleton och Sireci hade 1990 funnit hela 57 olika definitioner. Denna mångfald speglar en stor variation av bedömningsmodeller, som också ofta skiljer sig åt i sina filosofiska grundantaganden. Gipps (1994) menar

²² Ordet ”prestation” refererar i dessa sammanhang oftast till det svar, handling eller produkt, i vilken den kunskap som ska bedömas antas visa sig, och på basis av vilken denna kunskap bedöms. Det förekommer dock även att begreppet prestation används för att beteckna resultatet av sådan bedömning, exempelvis i form av ”poäng”. Eftersom det i det här sammanhanget är av viss betydelse att skilja dessa båda företeelser åt, så kommer jag i fortsättningen att använda begreppet prestation för att beteckna det som eleven har åstadkommit och som ska bedömas, och ”resultat”, ”poäng”, ”betyg”, eller ”omdöme” då det är någon av dessa företeelser som avses.

att det enda som egentligen förenar de bedömningsmodeller som kallas kriterierelaterade är att de inte är normrelaterade.

I Pophams definition av kriterierelaterad bedömning²³ beskrivs syftet som att

möjliggöra tolkning av studerandes prestationer på prov i relation till en uppsättning väl definierade kompetenser. (Hambleton & Sireci, 1997, s. 113)

Kriterierelaterad bedömning av elever eller studenter är ofta kopplad till formativa syften (se tidigare beskrivning av Catherine Taylors ”standardmodell”), men kan också syfta till att fastställa deras kompetens och används exempelvis för att avgöra om sökande till olika utbildningar är kvalificerade (behörighetsprov) och om personer, som har genomgått utbildningar av olika slag besitter det minimum av kompetens som fordras för att de ska kunna erhålla den yrkes- eller akademiska examen, som utbildningen syftar till och som kan ge tillträde till vissa yrken (Rowntree, 1977). Kriterierelaterad bedömning görs också för att kontrollera i vilken utsträckning, som elever har uppnått uppställda undervisningsmål och som grund för betygssättning,

Om den kriterierelaterade bedömningen inte är kopplad till sortering av eleverna är det i princip möjligt både att samtliga av dem som gjort provet godkänns och att samtliga underkänns. Om samtliga elever vid slutet av en kurs visar sig motsvara de kunskapskrav som ställs, skulle det ses som ett mycket gott resultat.

Normrelaterad bedömning

Vid norm- (eller grupp-)relaterad bedömning värderas elevernas kunskaper eller prestationer i förhållande till prestationerna hos de övriga som har gjort provet²⁴. Normrelaterad bedömning

²³ Enligt Hambleton och Sireci (1997) är denna den vanligaste definitionen av begreppet

²⁴ I Sverige kallades det normrelaterade betygssystem som erasttes i och med LPO 94 och LPF 94 även för *relativt*.

syftar alltså inte till att fastställa vad en elev kan, utan enbart till hur väl hennes eller hans prestation hävdar sig i relation till kamraternas.

Normrelaterad bedömning, som alltså inte ger någon egentlig information om vad den som bedömts kan, syftar till differentiering och rangordning och är därför främst användbart för urval. Vid normrelaterad bedömning av ett prov fastställs meritvärdet av elevernas prestationer på provet. Meritvärdet är ju beroende av hur väl de andra som gjort provet har presterat. Om ett prov ska ge underlag rangordning eller urval av eleverna, väljs bedömningskriterierna, liksom uppgifterna, därför på ett sådant sätt, att en viss spridning av eleverna åstadkoms. Man talar i sådana fall om ”differentierande” eller ”diskriminerande”²⁵ uppgifter. Om någon uppgift klaras av alla med godkänt resultat, så ses det inte som något positivt. Uppgifter, som inte åstadkommer någon spridning mellan eleverna är inte önskvärda i prov, som ska ligga till grund för normrelaterad bedömning.

Viktigt att notera är, att kriterier används även vid normrelaterad bedömning, om än i en annan mening. När elevernas svar bedöms, så sker det utifrån de kriterier som har satts upp för att svaret exempelvis ska godkännas. Dessa kriterier kan bestämmas på förhand, eftersom de hänför sig till ämnet och till undervisningsmålen. Huruvida ett svar ska betecknas som ”rätt” eller ”fel” kan ju inte vara beroende av hur många som har svarat på olika sätt. Först i ett andra skede av bedömningsprocessen, nämligen då värdet av en viss poängsumma ska bestämmas för att provet ska kunna användas som urvalsinstrument, normeras gruppens resultat, vilket innebär att varje elevs poängsumma, kallad ”råpoäng”, transformeras till ett värde på en given skala (exempelvis på det svenska högskoleprovet mellan 0,0 och 2,0).

²⁵ Begreppet ”diskriminerande” refererar i detta sammanhang enbart till provets egenskap att åstadkomma en spridning av eleverna utefter betygsskalan, att skilja ”dem som kan” från ”dem som inte kan” och antyder inte, vilket man ju med utgångspunkt i de sammanhang, som vanligen förknippas med ordet (könsdiskriminering och etnisk diskriminering) skulle kunna tro, att detta åtskiljande sker på grunder som inte är legitima.

Detta värde talar om hur nära gruppens genomsnitt, som respektive elevs prestation ligger (Rowntree, 1977).

Som exempel på bedömning, som på detta sätt involverar både kriterier och normer, kan nämnas IQ-testning. Rowntree skriver:

När vi använder standardiserade tester för att uppskatta intelligensen hos en grupp barn, utför vi alltså faktiskt en kriterierelaterad bedömning, eftersom barnens IQ faktiskt inte alls har något samband med kamraternas prestationer, utan bara med hur deras egen prestation är i jämförelse med de kriterier, som tagits fram genom att man utvärderat testen på en större population. Men även om testet är kriterierelaterat för provgruppen, är det inte det för populationen: testet är konstruerat på så sätt att de poäng, som uppnås av alla barn, skulle ungefär beteckna den s.k. normalfördelningen eller den klockformiga kurvan, där de flesta barn får medelpoäng och allt färre får högsta eller lägsta poäng. (Rowntree, 1977, s. 175)

Här föregås alltså den kriterierelaterade bedömningen av svaren av ett normrefererat urval av frågorna.

Analogt med detta (och baserat på en objektivistisk kunskapsteori, samt en föreställning om att potentialen för lärande är normalfördelad) anser Hambleton & Sireci (1997), liksom Sax (1989) att viss normativ information är nödvändig även vid vanliga "klassrumsprov" – detta för att få en uppfattning om vilka krav som är rimliga att ställa på eleverna. Om provet är "för lätt", så att alla lyckas, menar Sax, så kan det leda till att föräldrar tror att deras barn är mera högpresterande än vad som faktiskt är fallet. Om provet är "för svårt" å andra sidan, så att ingen lyckas, får det negativa konsekvenser för de enskilda elevernas motivation och för klassens arbetsmoral. Dessa synpunkter kan ses som uttryck för en utbildningssyn som bygger på uppfattningen att de riktigt goda inlärningsresultaten i skolan bara kan uppnås av ett fåtal elever, vilket i sin tur bygger på teorin om intelligens som konstant och normalfördelad.

Då normrelaterad bedömning ofta är kopplad till beslut som har viss (eller stor) betydelse för elevernas framtida livschanser, tillmäts reliabiliteten stor vikt. Av denna anledning har prov, som ska ligga till grund för sådan bedömning ofta svarsformat, som innebär ett begränsat antal väldefinierade svarsmöjligheter, där endast en representerar ett acceptabelt svar.

Som påpekades i den första delen av detta arbete, utgjorde de normrelaterade proven ur historiskt perspektiv en viktig del i ”det moderna projektet” (och av det som har kallats den ”sociala ingensjörskonsten”) som medel för att likställa livschanserna för individer med olika ursprung. Att intresset för kriterierefererade bedömningsmodeller sedan ett par decennier har vuxit bland lärare, forskare och utbildningsplanerare, kan förstås som en del av den process där bedömningsformerna ska anpassas till de nya villkor, som övergången från det moderna till det postmoderna samhället ställer på utbildningssystemet. Denna utveckling har ju, som beskrevs i avsnittet om ”det nya provparadigmet” bl a inneburit ett ökat behov av bedömningsformer, som motiverar snarare än slår ut svagpresterande elever, är förankrade i sin lokala kontext, som ger läraren tydlig information om de enskilda elevernas kunskaper, progression och behov och som kan ge stat och kommun underlag för bedömning av utbildningskvaliteten vid olika skolor. Det finns också behov från statens sida av metoder för att bevaka den reella existensen av en nationell läroplan, dvs att kontrollera att undervisningen runtom i landet, trots regionala skillnader vad gäller både form och innehåll, ändå har en gemensam kärna av kunskaper (Broadfoot, 1995).

Att bedöma sammansatta kunskaper med god tillförlitlighet

I ett försök att kombinera de traditionella, objektiva, normrelaterade provens goda mättekniska egenskaper med de lokalt utformade, kriterierelaterade kvalitativa provmodellernas fördelar vad gäller pedagogiska egenskaper och innehållslig relevans, så har man i Sverige på senare år försökt att utveckla nationella prov med uppgifter, som fordrar att eleverna resonerar och själva konstruerar sina svar, och gemensamma bedömningskriterier som omfattar även högre kognitionsnivåer. Det finns också en tydlig intention att betygskriterierna, som hör till de svenska läroplanernas kursplaner, ska användas för att utveckla såväl som synliggöra och bedöma elevernas lärande i flera dimensioner. Skolverket skriver i sitt kommentarmaterial ”Betyg och betygssättning” (2001):

Det är också viktigt att eleverna som ett led i sitt lärande uppmuntras till och övas att själva bedöma sina kunskaper och sitt lärande, och också att bedöma värdet av och meningen med kunskaperna. (...) Bedömning handlar således i dagens svenska läroplaner om något mer än att rangordna elever efter poäng på olika prov. Den syftar till att kunna identifiera enskilda levers kunskapskvaliteter för att på bästa sätt kunna beskriva starka och svaga sidor. Bedömningen syftar vidare till att så tydligt som möjligt ge läraren underlag för att kunna informera elever och vårdnadshavare om hur det går för eleven i skolan och till att ge ett bra underlag för att planera för elevernas fortsatta utveckling. (ibid., s. 17)

Vidare postuleras att ”elevernas prestationer ska sättas i relation till mål och kriterier beskrivna i olika kunskapskvaliteter” (s.18), och kursplanerna innehåller som stöd för betygssättningen ämnesspecifika kriterier för varje betygssteg. I dessa ingår flera olika aspekter på kunskap, vilket innebär att eleverna för att få ett högt betyg måste kunna demonstrera, inte bara att de besitter en stor mängd faktakunskaper, utan också att de t ex har en förståelse för sammanhang och kan använda sina kunskaper (Skolverket, 2001). Vid betygssättningen ska läraren (enligt Grundskoleförordningen 7 kap.) ”utnyttja all tillgänglig information om elevernas kunskaper i relation till kursplanen”.

Målen som anges i kursplanen för vad eleverna ska ha uppnått i slutet nionde skolåret i ämnet samhällskunskap får tjäna som ett exempel på vilka slags kunskaper som enligt den svenska läroplanen ska fokuseras i bedömningen.

Enligt dessa skall:

- känna till grundläggande mänskliga rättigheter och skyldigheter,
- kunna reflektera över människors villkor och hur dessa skiftar utifrån kön och social och kulturell bakgrund,
- förstå de gemensamma och grundläggande demokratiska värden som vårt samhälle vilar på samt kunna tillämpa ett demokratiskt arbets- och beslutssätt,
- ha grundläggande kunskaper om det svenska styrelsesättets framväxt och samhällssystemet uppbyggnad samt förstå innebörden av det grundläggande norm- och rättssystemet i Sverige,
- ha kunskaper om internationella förhållanden samt kunna diskutera relationer och samarbete i ett globalt perspektiv,
- ha kännedom om samhällsekonomi, hushållsekonomi och privatekonomi och kunna diskutera möjliga vägar som leder mot ett hållbart samhälle såväl lokalt som globalt,

- kunna använda olika kunskapskällor samt kunna sammanställa, bearbeta, granska och värdera uppgifter från olika källor och opinionsbildare samt redovisa resultatet

För de olika betygen krävs sedan att eleverna visar sin kompetens inom olika aspekter av ämnet. För att bli godkänd fordras t ex att eleven när det gäller *källkritisk förmåga* ska ”kunna använda olika kunskapskällor samt kunna sammanställa, bearbeta, granska och värdera uppgifter från olika källor och opinionsbildare samt redovisa resultatet”. För att få betyget ”Väl godkänd” fordras att eleven ”samlar in fakta om samhällsfrågor från olika kunskapskällor och värderar och drar slutsatser av dessa fakta vid en sammanställning”, och för betyget ”Mycket väl godkänd” att hon ”använder och sammanställer fakta från olika källor, bedömer deras relevans, granskar uppgifterna kritiskt samt skiljer på fakta och värderingar”.

Ett amerikanskt exempel på hur man har försökt att finna standardiserade kriterier för att bedöma mera sammansatta kunskaper, utgör den amerikanska staten Kentuckys bedömningsmatris för betygssättning av elevernas prestationer på gemensamma tillämpningsuppgifter, som här beskrivs utifrån Cunningham (1998, s. 134).

AVANCERAD	Eleven fullföljer alla viktiga delar av uppgiften och kommunicerar sina idéer tydligt. Eleven visar att han eller hon förstår aktuella begrepp och eller processer på ett djupare plan (kan generalisera, tillämpa och se analogier).
KVALIFICERAD	Eleven fullföljer de flesta av uppgiftens viktiga delar och kommunicerar sina idéer tydligt. Eleven visar att han eller hon förstår de centrala begreppen även om han eller hon bortser från eller missförstår mindre viktiga idéer eller detaljer.
LÄRLING	Eleven fullföljer några av uppgiftens viktigare delar kommunicerar dessa idéer tydligt. Elevens svar visar att det finns luckor i hans eller hennes begreppsliga förståelse.
NOVIS	Eleven klarar inte av att skapa ett sammanhang, eller presentationen innehåller bara återgiven information. Svaren är helt fel eller irrelevanta.

TABELL 3. Bedömningsmatris från den amerikanska delstaten Kentucky. Efter Cunningham (1998).

Broadfoot (1995) tar upp en rad exempel på nationella program för klassumsbaserad bedömning, som har öppnare svarsformat och är inriktade mot mera komplexa färdigheter än vad som tidigare varit vanligt vid storskaliga bedömningar. Sådana nationella program finns exempelvis i Storbritannien, ett par delstater i Australien, i Nya Zeeland och i Frankrike.

Ett äldre och välkänt exempel på en bedömningsmatris med syftet att ge lärare universella kriterier för att kunna göra mera tillförlitliga bedömningar av sammansatta kunskaper utgör SOLO-taxomin²⁶, som konstruerades av de båda australiensiska forskarna Biggs och Collis (1982).

Principen för SOLO-taxomin är att resultatet av elevernas behållning av undervisning av sluten natur, dvs undervisning med ett entydigt och på förhand specificerat mål, bedöms på basis av den strukturella komplexiteten i elevens svar på en provfråga, eller annan prestation, som är avsedd att testa elevens kunskap. Taxomin består av fem nivåer. Varje nivå ges en siffra mellan 1 och 5 och "övergångsnivåerna" däremellan betecknas med siffran för respektive nivå följt av ett A, t ex 2A, som betecknar ett "bättre" svar än 2. De olika nivåerna härstammar, enligt författarna, från Piagets kognitiva utvecklingsstadier, men följer inte dennes teorier om tankeförmågans åldersbundenhet. I tabell 2 presenteras de fem nivåerna.

²⁶ SOLO står för the structure of the observed outcome

PRESTRUKTURELL

- 1) Eleven undviker frågan eller upprepar innehållet i den.
- 1a) Eleven försöker besvara frågan, men lyckas bara delvis få med något signifikant.

UNISTRUKTURELL

- 2) Svaret baserar sig uteslutande på en aspekt av resonemanget ifråga och drar en förenklad slutsats, som därför blir felaktig eller dogmatisk.
- 2a) Eleven försöker hantera två aspekter, men de är inte kongruenta och hon lyckas därför inte dra någon slutsats.

MULTISTRUKTURELL

- 3) Flera konsistenta aspekter av data presenteras, men konflikter ignoreras och en förenklad slutsats dras på ett urval av endast samstämmig information.
- 3a) Eleven uppmärksamma inkonsistenser och konflikter, men kan inte få ihop en övergripande förklaring.

RELATIONELL

- 4) Hela eller huvuddelen av resonemanget finns med. Eleven försöker få ihop en helhetsbild, som också gör motsägande data begripliga.
- 4a) Antydningar om att slutsatsen som har dragits är en av flera möjliga.

UTVIDGAT ABSTRAKT

- 5) Eleven uttrycker en medvetenhet om att det givna exemplet är ett fall av en mera generell och övergripande princip. Flera slutsatser hålls öppna.
-

TABELL 4. SOLO-taxomin. Efter Biggs & Collis (1982)

Biggs och Collis (ibid.) poängterar att ”det observerade utfallet” av elevernas lärande, i praktiken ofta skriftliga svar på provfrågor, endast i optimala fall motsvarar elevernas kompetens – oftast innehåller provsituationen faktorer, som gör att det observerade utfallet av elevers lärande är mindre avancerat, än det faktiska. Exempel på sådana faktorer är elevens motivation, nervositet och skriftspråklig förmåga. För att undvika de validitetsproblem i bedömningen som sådana störande faktorer medför, begränsar Biggs och Collis slutsatserna av bedömningen till att gälla enbart ”det observerade utfallet”.

Global och analytisk bedömning

När bedömningen av ett arbete, exempelvis en uppsats, sker på basis av bedömarens helhetsintryck, så kallas bedömningen "holistisk", eller "global". Hos Scriven (1996), innebär globala bedömningar att hela processen bedöms. Det innebär att om bedömningen enbart riktar sig mot resultatet, så kan den inte betecknas som global. Sådana bedömningar, som gäller en eller flera delar av processen benämns i stället analytiska (ibid.).

Analytisk utvärdering kan ske i förhållande till antingen olika dimensioner (om vi tar bedömningen av en uppsats som exempel, så kan sådana dimensioner vara sakkunskap, analys, originalitet, språkbehandling, struktur och koncentration) eller referera till olika delar (i en uppsats till exempel inledning, metodkapitel, resultat, diskussion). Det förra brukar kallas dimensionell utvärdering och det andra komponentutvärdering eller atomistisk utvärdering.

Den kriterierelaterade bedömningens kunskapsteori

I mångfalden av kriterierefererade bedömningsmodeller kan en uppdelning i två huvudinriktningar urskiljas. Denna uppdelning kan jämföras med samhälls- och beteendevetenskapens omdebatterade, men, som det tycks, ändå ganska seglivade uppsplätning i kvalitativa och kvantitativa forskningsansatser. Lärares arbete med att producera och tolka data om elevers kunskaper och lärande för att kunna utvärdera och individualisera undervisningen och för att kunna sätta betyg på eleverna är ju också i många stycken artligt forskarens arbete – det kan ses som en fråga om hur noggrant, systematiskt och nyanserat som detta arbete genomförs, dokumenteras och analyseras.

De bedömningsmodeller, som skulle kunna beskrivas som präglade av naturvetenskapliga forskningsideal utgår då, liksom "kvantitativ forskning", från ett deduktivt kunskapssökande, medan de konstnärligt inriktade bedömningsmodellerna utgår

från den ”kvalitativa forskningens” induktiva väg till kunskap (se t ex Hartman, 1998).

Naturvetenskapligt orienterad bedömning

De ”naturvetenskapliga” bedömningsmodellerna innebär sålunda, att elevernas prestationer bedöms utifrån på förhand specificerade, för eleverna bekanta, och noggrant definierade kriterier. Precis som vid traditionell, normrelaterad bedömning fästs stor vikt vid jämförbarhet, objektivitet och rättvisa. För att det ska vara möjligt att specificera bedömningskriterierna på förhand, så måste eleverna i början av ”kursen” ha exakt information om vad de förväntas att lära sig. Som exempel ser Cunningham (1998) det som en absolut nödvändighet, att eleverna har fullständig information om de kriterier, som ska tillämpas vid bedömningen av deras arbete, d v s att bedömningen är ”transparent”. Särskilt vid bedömningar, som är av uppenbar betydelse för elevernas framtid är detta viktigt, då det handlar om deras rättssäkerhet.

Eisner (1991:170 f) beskriver fyra problem som är förknippade med kravet på att bedömningskriterierna ska vara entydigt formulerade redan vid undervisningens start, så att eleverna vet vad som förväntas av dem och bedömningen kommer att rikta sig uteslutande mot det som undervisningen faktiskt har handlat om:

- 1) A. Kunskapskvaliteter behandlas, eftersom man betraktar allt som kausala samband, som instrumentella. B. Det specifika går förlorat – inga mätningar kan ge svar på vad det faktiskt är som händer i en specifik klass med en specifik lärare. Det specifika blir istället till ”brus” eller mätstörningar, då det som inte är definierat på förhand faller utanför.
- 2) Det leder till standardiserad mål-medel-inriktad undervisning, och en reduktionistisk syn på läraryrket. Om lärare skulle arbeta enbart utifrån på förhand uppställda mål, skulle det bli en mekanisk instruktionsboksliknande undervisning. Koncentrationen på att uppnå mål i framtiden skulle dessutom ske på bekostnad av lärarens närvaro i nuet. Lärare

måste under sin arbetsdag fatta hundratals snabba beslut baserade på intuitiva sammanvägningar av en mängd faktorer på olika nivåer, poängterar Eisner och menar, att det som utmärker en skicklig lärare är kvaliteten i sådana ”i-flykten-beslut”. Kraven på i förväg specificerade kriterier, innebär att sådana aspekter av lärarkompetensen som förmågan att snabbt tolka och fatta beslut i komplexa situationer, inte kan utvärderas. Därmed, menar Eisner, blir de också mindre rationella, än de ger sig ut för att vara.

- 3) Objektiviteten som norm är uttryck för en förlegad epistemologi, som innebär att världen ska beskrivas så som den verkligen är – kunskapen växer allteftersom vi upptäcker hur världen ser ut. Med objektiviteten som kunskapsideal bortser man också från flera viktiga aspekter av människors förhållande till världen: *”intimation, metaphor, analogy, poetic insight have little place in such a view”*.
- 4) Det begränsar lärares möjligheter att lära sig om elevernas tänkande, kunskaper, utveckling.

Humanistiskt eller konstnärligt orienterad bedömning

Den humanistiskt orienterade bedömningsmodellen innebär, att kriterierna växer fram hos bedömaren i förhållande till det unika i det som bedöms. På ett abstrakt plan kan kriterierna definieras på förhand, men den konkreta innebörden tar form i bedömarens subjektiva tolkningsprocess, där såväl ”tyst” som manifest kunskap medverkar (exempelvis bedömarens epistemologiska och ideologiska grund, ämneskunskaper och erfarenheter av tidigare arbeten av motsvarande art). Detta hermeneutiska förhållningssätt till bedömning och till bedömarens roll avviker starkt från den vetenskapliga modellens normer. Eisner (ibid.) använder två begrepp från konstvetenskapliga sammanhang – ”konnässörskap” och ”kritik” – för att gestalta detta förhållningssätt. Han beskriver ”konnässören” på följande sätt:

The connoisseur is someone who knows what he or she is looking at, or listening to, or reading. The connoisseur is someone whose ability to noti-

ce the subtleties that count in some domain, to experience nuance, to recognize the import of a painting, a poem or a symphony is highly refined. It requires an active intelligence and the application of refined schemata. A connoisseur is someone who has worked at the business of learning how to see, to hear, to read the image or text, and who, as a result, can experience more of the work's qualities than most of us. (Eisner, 1991, s. 174)

För att resultatet av "konnässörens" bedömning ska kunna förmedlas till andra, ikläder sig denne rollen som "kritiker", d v s en expert som hjälper dem, som saknar kunskaper och erfarenheter inom området, att se dessa sublimes kvaliteter, undermeningar eller nyanser. Detta sker ofta med hjälp av det skrivna ordet, och Eisner ser bedömning i pedagogiska sammanhang idealt som artiklens verksamhet, som konstkritiker eller filmkritiker sysslar med när de bedömer och skriver om konst eller film.

Bedömarens uppgift är, enligt Eisner, alltså, att i kraft av sin expertkunskap kartlägga och "lägga ut texten" om arbetets kvaliteter och brister. För att kunna göra detta krävs att bedömaren, exempelvis läraren, besitter stor känslighet för detaljer, "ett tillräckligt differentierat spektrum av scheman för att kunna uppleva och förstå en liten bit av världen", och dessutom förmågan att skriva på ett sätt som gör att läsaren får en nyanserad och tydlig bild av de komplexa fenomen och processer, som lärande och undervisning består av (s. 183 ff).

"Connoisseurship" och "criticism" kan också sägas representera en form av kunskapssökande, som ligger mycket nära etnografins, en inriktning där ju forskarens erfarenheter och inlevelseförmåga är helt centrala och inte kan ersättas av standardiserade regler och procedurer och där sättet att kommunicera den kunskap som vunnits tillmätts mycket stor betydelse.

Det kvalitativa forskningsparadigmets prioritering av kontext, mening, autenticitet och kommunicerbarhet framför reliabilitet, representativitet och statistisk generaliserbarhet, har alltså en motsvarighet i detta förhållningssätt till bedömning. Samma principer är också bärande i exempelvis olika slag av tillämpningsprov och portföljer.

När det gäller bedömningen av högskolestudenter och studenter vid yrkesutbildningar av olika slag, där en väsentlig del av den kompetens som åsyftas består av så kallad ”tyst kunskap” och ”kunskap i handling”, är konflikten mellan de krav som studenternas rättssäkerhet ställer på entydiga och i förväg uppsatta kriterier och de krav på validitet och formativa egenskaper, som upprätthållandet av en hög utbildningsstandard ställer, särskilt framträdande. Exempelvis menar Lauvås (1998), att om ett omdöme om en högskolestudent ska upplevas som välgrundat, objektivt och rättvist fordras ofta att examinatorn kan redovisa de kriterier utifrån vilka bedömningen gjorts – trots att detta inte alltid låter sig göra. Att en examinator inte kan redovisa kriterierna betyder dock inte nödvändigtvis att kriterierna är dåliga, utan de kan vara både precisa och relevanta – Lauvås jämför med muraren, som utan att kunna formulera på vilka grunder han gör sin bedömning, ändå kan avgöra om ett murararbete är fackmannamässigt utfört. När det gäller företeelser som ”den vetenskapliga hållningen” hos studenter, kan det samtidigt vara så, att två olika examinatore bedömer en student olika, beroende på att de lägger olika innebörd i begreppet.

Hur kan resultaten rapporteras?

Sättet att rapportera resultatet av bedömning av elevers kunskaper väljs, precis som provmodellen som sådant, utifrån syftet: formativa syften kräver utförlig kvalitativ och konstruktiv feedback (Shepard, 2000). Resultatet av summativa bedömningar kan rapporteras antingen i form av poäng, eller rangordning (som siffror, bokstäver, eller vissa standardiserade omdömen), eller i form av kvalitativa utsagor som beskriver vad eleven kan och inte kan i relevanta avseenden. Omdömet eller betyget avser hur en elev har presterat vid ett visst tillfälle, t ex ett prov, eller elevens kompetens mera allmänt och då grunda sig på information från flera bedömningstillfällen (jämför betygsgrundande bedömning i svensk grund- och gymnasieskola, där betyget ska grundas på en bedömning av *all* tillgänglig information om elevens kunskande i ämnet).

I USA har man i vissa stater delvis ersatt de traditionella sätten att redovisa elevernas kunskapsnivå med vad man kallar "behavioural mastery reports" (Sax, 1989) (jämför det tidigare nämnda svenska *Läs- och Skrivutvecklingsschemat*, LUS), en sorts formulär med deskriptiva kategorier för olika nivåer i ämnet, där det framgår vad en elev kan. Dessa anses bli ha de stora fördelarna att de ger föräldrarna en god bild av elevernas kunskaper, gör eleverna mindre konkurrensinriktade och är värdefulla för läraren i planeringen av det pedagogiska arbetet (ibid.). Försök med att ersätta de abstrakta siffer- och bokstavsbedömningsbetygen med mera detaljerade beskrivningar har även gjorts i Australien, där jämförbarhet när det gäller elevernas kunskaper över landet åstadkoms, inte genom att man ger prov med samma uppgifter, utan genom att man tillhandahåller exempel på de olika betygsnivåerna. Ämnena delas upp i flera aspekter som bedöms och redovisas var för sig (Broadfoot, 1995). I Sverige ges elever och föräldrar kvalitativa beskrivningar av elevernas kunskaper vid muntliga "utvecklingssamtal".

Sifferomdömen kan visserligen vara nödvändiga i vissa sammanhang, men Wolf (1994) menar att de pedagogiska fördelar som är förknippade med de kvalitativa omdömena, och det faktum att sifferbetyg och standardiserade omdömen ofta i hög grad är beroende av opövande antaganden, gör att man ändå bör undvika att rapportera resultatet i form av siffror och andra abstrakta symboler, om det är möjligt att beskriva vad eleverna har lärt sig i mera specifika och konkreta termer.

Läsfärdighet i olika åldrar, skulle exempelvis kunna beskrivas i termer av vilken typ av texter som eleven kan förstå, elevens ordförråd och syntaktiska förmåga, skriver Wolf (ibid.). Ett annat exempel är bedömningen av uppsatser, där Wolf rekommenderar att läraren redovisar vilka kvaliteter som bedöms, t ex tydlighet, och exemplifierar varje nivå så att det blir tydligt för elever och föräldrar vad som värderas. Istället för att ange ett enda medelvärde för studenternas resultat, rekommenderar Wolf istället att läraren anger hur eleverna presterade med avseende på alla de olika kriterierna.

Bedömningens kvalitet

”Mätmodellen” och ”standardmodellen”

Hur man kan avgöra om ett prov är bra (och därmed hur man kan göra bra prov) och hur man kan bedöma relationen mellan en elevs provresultat och de kunskaper som det är avsett att spegla är naturligtvis ytterst angelägna frågor för lärare. Dessa frågor handlar om kvalitet dels i proven som sådana, dels i den process, där elevernas prestationer tolkas och värderas.

Det finns två principiellt skilda sätt att betrakta sådana kvalitetsfrågor. De hänför sig till olika sätt att se på syftet med bedömning och, i förlängningen, utbildning (Taylor, 1994).

Om syftet med bedömningen är att rangordna eleverna eller att bestämma värdet av en elevs prestation i förhållande till hur övriga elever har presterat, är t ex en viktig kvalitet hos provet att valet och sammansättningen av uppgifter är sådana att elevernas resultat i slutändan skiljer sig från varandra. Om tillräckligt många gör provet, ska utfallet idealt vara att resultaten blir normalfördelade. Frågan om ett prov är bra, kan ur detta perspektiv alltså inte besvaras förrän man ser hur utfallet blir – ett prov vars uppgifter exempelvis väl speglar uppsatta kunskapsmål och av eleverna uppfattas som motiverande, måste göras om ifall alltför många elever lyckas eller misslyckas (ibid.). Ett sätt att åstadkomma den avsedda spridningen av resultaten är att konstruera provet så att det innehåller en väl avvägd sammansättning av ”lätta” och ”svåra” uppgifter, d v s uppgifter som många respektive få elever förväntas klara av. Mängden av uppgifter har också betydelse – ett prov med många uppgifter ger i princip större spridning än ett med få, och risken att en elev lyckas eller misslyckas beroende på slumpen minskar.

Om provet ska testa elevernas kunskaper inom ett område ska frågorna utgöra ett stickprov av mera specifika kunskaper inom området – tanken är att den grad i vilken eleven besitter kun-

skaper om området som helhet, återspeglas i den grad hon klarar av att lösa ett urval av specifika frågor inom området. Stickprovet ska då vara representativt för området och okänt för eleverna fram till provtillfället så att de inte har möjlighet att förbereda sig extra för dessa uppgifter (ty annars blir ju resultatet på provet inte längre representativt för elevens kunskaper inom det testade området som helhet). När det gäller bedömningen av elevernas uppvisade kunskaper får den inte påverkas av vem som bedömer: resultatet ska enbart spegla elevens kunskaper och kriterierna för huruvida ett svar visar att eleven besitter den kunskap det gäller ska vara så entydiga som möjligt.

Om syftet med bedömningen å andra sidan är att fastställa vad eleverna kan och om eller i vilken utsträckning som de besitter de kunskaper som undervisningen syftar till, är det andra kvalitetskriterier som är relevanta. Uppgifterna måste då vara av sådant slag att varje elev får optimala förutsättningar att demonstrera de kunskaper som bedömningen gäller – variation i uppgiftsformat och möjlighet för eleverna att välja uppgiftsformat ökar dessa förutsättningar. En annan konsekvens är att det är rimligt bedömningen av varje elevs kunskaper sker när eleven är ”redo” för det, vilket innebär att olika elever bedöms vid olika tillfällen (Taylor, *ibid.*).

Kvalitetsbegrepp

I den vetenskapliga litteraturen används ofta begreppen reliabilitet och validitet när det handlar om att värdera kvaliteten i information och att värdera tolkningar och slutsatser av den information man har tillgång till. ”Reliabilitet” kan i detta sammanhang definieras som ”den utsträckning i vilken ett prov skulle generera samma eller liknande poäng om det gavs av två olika bedömare, eller om det gavs en andra gång till samma elev” (Gipps, 1994, s. 2), och validitet som ”ett integrerat värdeomdöme, som gäller den utsträckning i vilken empiriska belägg och teoretiska resonemang stödjer giltigheten och riktigheten hos slutsatser och handlingar som görs på basis av provresultat eller andra former av bedömningar” (Messick, 1993).

Båda begreppen har sitt ursprung i kvantitativ forskning, inom pedagogikämnet i "mätmodellen" och psykometrin. Medan validitetsbegreppet emellertid intar en central position i diskussionen om kvalitet även i "standardmodellen", d v s prov som syftar till att fastställa elevers kunskap i förhållande till olika mål, så är reliabilitetsbegreppet mindre tillämpligt i dessa sammanhang. Reliabilitetsmått av olika slag, som redovisas nedan, är istället avgörande för kvaliteten i "mätmodellen" – vilket ju är helt i konsekvens med dess syfte att resultera en sortering av eleverna, som är korrekt utifrån de grundantaganden man gör om lärande och utbildning.

Reliabilitet

Graden av samstämmighet mellan olika bedömare benämns interbedömarreliabilitet. Vid standardiserade prov testas ofta interbedömarreliabiliteten och gränser sätts upp för hur stor varians som tolereras. Stabilitet (variansen i en elevs resultat på samma uppgift över tid) och response consistency (variansen i en elevs resultat på olika uppgifter som är avsedda att testa samma sak) är två andra former av reliabilitet. Vid prov, som är har ett klart definierat svarsutrymme och innehåller ett stort antal uppgifter, kan reliabiliteten beräknas statistiskt (Cunningham, 1998).

Validitet

I det kvantitativa forskningsparadigm, där även validitetsbegreppet har sitt ursprung, brukar man tala om tre olika slag av validitet: Innehålls-, kriterie- och begreppsrelaterad validitet. Innehållsrelaterad validitet refererar i vilken utsträckning frågorna eller uppgifterna täcker undervisningsinnehållet eller undervisningsmålen. Den innehållsrelaterade validiteten kan referera antingen till kursplanemål eller andra typer av utbildningsmål, eller till den genomförda kursen, d v s den undervisning som eleverna faktiskt har haft. Enligt Yalow och Popham (i Messick, 1993) föreligger innehållslig validitet i den utsträckning som provet utgör ett representativt stickprov på de kunskaper och färdigheter som man är intresserad av att eleverna ska ha

utvecklat inom det aktuella området. ”När man testar en gymnasieelevs matematikkunskaper utgörs det aktuella området inte av det som undervisningen råkar ha handlat om, utan av det som borde kunnas” (ibid. s. 68).

Ett sätt att försäkra sig om god innehållslig validitet i sådan bedömning som avser elevernas kompetens i förhållande till ett viss ämnesområde är, att involvera ämnesexperter med god kännedom om aktuella läroplaner och läromedel i utformningen av uppgifter och bedömningskriterier²⁷. Det kallas att göra en expertvalidering (ibid.).

Kriterierelaterad validitet refererar till bedömningskriteriernas relevans i förhållande till undervisningsmålen och begreppsvaliditet³¹ till relationen mellan det som faktiskt mäts och det som man drar slutsatser om (Cunningham, 1998).

Uppspaltningen i innehålls-, begrepps- och kriterierelaterad validitet utgår emellertid från de principer, som utmärker ”konventionella prov”, nämligen dels att bedömningen sker indirekt, dvs att istället för att bedöma den aktuella kompetensen direkt, så drar man slutsatser om den på basis av hur eleverna har svarat på vissa frågor, dels bedömningen görs på basis av någon form av stickprov. När det gäller ”tillämpningsprov” eller autentiska prov är detta inte relevant, eftersom de ju bygger på principen att den kompetens som står i fokus bedöms direkt på basis av hur framgångsrikt eleven förmår att genomföra sådana uppgifter, som fordrar den aktuella kompetensen mobiliseras.

I bedömningsmodeller som syftar till att fastställa vad eleverna kan (och till att stödja deras fortsatta lärande), ingår ofta en värdering av hur provet som sådant påverkar undervisningen och elevernas lärande. Moss (1994) poängterar särskilt, att i den

²⁷ Begreppsvaliditet är den svenska ord som vanligen används som översättning för det engelska begreppet *construct validity*. Enligt Cunningham (1998, s. 39) betecknar emellertid ordet ”construct” en uppsättning av beteenden. Med denna innebörd skulle *construct related validity* beteckna relationen mellan det beteende som man faktiskt studerar och det som man drar slutsatser om.

mån som tolkningarna av resultaten har sociala konsekvenser, så ingår det i valideringen att utvärdera även dessa. Om provet skulle visa sig ha oönskade sociala konsekvenser (exempelvis, att en avsevärt större andel flickor än pojkar underkändes och därmed utestängdes från vissa fortsatta utbildningsmöjligheter), så innebär detta inte i sig att provet har låg validitet, menar Messick (1993). Däremot ingår det i valideringen av provet, att försäkra sig om att dessa oönskade konsekvenser inte beror på att några andra faktorer, än dem som provet är avsett att mäta, har påverkat resultatet.

Att valideringen omfattar konsekvenserna, inte bara av hur uppgifterna och resultatet tolkas, utan också av hur provet påverkar undervisningen (lärarens val av stoff och metoder, och hur undervisningen organiseras) och elevernas lärande (kunskaps-syn, lärandestrategier, förståelse av ämnet etc.) är ofta självklart inom "det nya provparadigmet", då ju former som "tillämpningsprov", "portfolios" och autentiska prov har utvecklats just med syftet att förbättra elevernas lärande²⁸. Den aspekt av validiteten, som handlar om exempelvis sociala och pedagogiska konsekvenser av bedömningen, brukar benämnas konsekvensvaliditet (t ex Moss, 1995).

Både Birenbaum (1996) och Messick (1993) påpekar att det som ska valideras är inte är provet eller bedömningsproceduren i sig utan de slutsatser som dras på basis av provpoäng eller andra indikatorer, d v s "slutsatser om vad poängen står för och om vilka praktiska implikationer som denna tolkning innebär" (Messick, 1993: s. 13). En viktig poäng här är att validiteten ses som kontextberoende. Det betyder att en förutsättning för att man ska kunna dra giltiga slutsatser är att provuppgifterna är meningsfulla för *eleven i det sammanhang som de ges*. Raven (1991) påpekar att elevers förmåga bara kan bedömas i relation till mål som de själva värdesätter, eftersom viktiga förmågor bara kommer till uttryck när man arbetar mot mål som upplevs

²⁸ Jämför med aktionsforskning, där ett centralt validitetskriterium är att den förändring, som man har arbetat rör att åstadkomma, vid en uppföljande utvärdering visar sig faktiskt ha ägt rum.

som angelägna att uppnå. Bristande motivation hos eleven kan alltså försvaga sambandet mellan provprestationen och den kunskap, som denna antas indikera. Men det handlar också om hur eleven förstår sammanhanget. Messick (1993, s.14) exemplifierar med, att ett barn kanske inte tycks förstå prepositionen *på* i ett sammanhang, där han ombeds rita en cirkel på en fyrkant, men däremot när han ombeds ställa en kloss på ett en leksakslastbil. Ett annat klassiskt exempel på hur barns möjligheter att demonstrera sin kunskap beror på barnets relation till det sammanhang, i vilket kunskapen testas, utgör ett i Donaldson (1979) återgivet experiment, som Piaget gjorde för att testa barns förmåga att inta en annan persons perspektiv (d v s *decentrera*). Donaldson visade, att barn, som enligt Piagets test ännu inte hade utvecklat denna förmåga, kunde klara en uppgift, som krävde just denna förmåga, men som uppfattades som meningsfull av barnen, om den var utformad som en leksituation som de kände igen. Ett annat sätt att uttrycka detta är att det ur validitetssynpunkt är viktigt att specificera det sammanhang där resultatet eller slutsatserna gäller, d v s att man anger gränserna för hur långt de slutsatser som man har dragit kan generaliseras. Det betyder inte att ett resultat måste vara universellt giltigt, utan bara att validiteten förutsätter att slutsatserna begränsas till att gälla i en specifik kontext. Om man av iakttagelsen att barnet inte kan demonstrera sin förståelse av prepositionen *på* i ett visst sammanhang, drar slutsatsen att barnet inte alls förstår begreppet, så har man gjort en alltför långtgående generalisering. Ett sätt att söka gränserna för den kontext, som resultatet kan generaliseras till är, att testa barnets förståelse av samma begrepp i ett annat sammanhang.

Relationen mellan reliabilitet och validitet i bedömningen

Stora möjligheter för eleverna att välja uppgiftsformat och innehåll i proven har alltså visat sig gynna validiteten i den mening att provresultatet ger en så rättvisande bild av elevens faktiska kunskaper som möjligt. Men det har också visat sig att ju större sådana valmöjligheter som provet innebär, desto svårare är det att uppnå hög interbedömarreliabilitet (Birenbaum, 1996). Vid autentiska prov och tillämpningsprov har det dessutom visat

sig att variansen till följd av att elever presterar olika beroende på uppgiften t o m brukar vara större än variansen mellan olika bedömare (ibid.). Detta visar att elevernas förmåga är i hög grad kontextberoende. Här finns alltså en paradox: om man utformar prov så att de tar hänsyn till att elevernas förmåga är kontextberoende, så påverkar det stabiliteten och därmed reliabiliteten, negativt.

Tillämpningsprov och andra former av "alternativa prov" innehåller också ofta ett fåtal uppgifter (ibland bara en). De statistiska metoder som finns för att beräkna reliabiliteten syftar till att fastställa stabiliteten i elevernas svarsmönster – ju stabilare svarsmönster, desto högre reliabilitet. Om provet innehåller få frågor fungerar inte dessa metoder (Taylor, 1994).

Vad innebär då dessa motsättningar mellan validitet och reliabilitet i praktiken? Hur kan de hanteras?

Exempelvis Cunningham (1998) menar, att tillämpningsprov bör undvikas i sammanhang där bedömningen ska ligga till grund för viktiga beslut (exempelvis urval och examinationer) av just det skälet att möjligheterna att veta om provet skulle ge samma resultat vid nästa tillfälle är kraftigt begränsade. I många amerikanska stater, men också i Sverige, har man dock valt att försöka utveckla "alternativa" provformat även i så kritiska sammanhang som storskaliga utvärderingsprogram och i det nationella examinationssystemet. I USA har det emellertid ofta inneburit att kriterierelaterade bedömningsmodeller och tillämpningsprov av olika slag har passats in i mätmodellen med dess bakomligganden syften och värderingar och dess underliggande syften och värderingar, liksom dess procedurer för att säkra validitet och reliabilitet (Taylor, 1994). Konsekvenserna av detta för eleverna och undervisningen blir då desamma som för normrelaterade prov, menar Taylor. Hon anser att det verkliga problemet inte handlar om att reliabiliteten inte går att säkerställa om man använder vissa provformat, utan om att en och samma bedömningsmodell inte kan användas för att både ge en sann och fullödlig bild av elevernas kunskaper eller skolornas måluppfyllelse och sortera differentiera och rangordna elever eller skolor.

Alternativa sätt att resonera om kvalitet i bedömning

Om man vill dra nytta av de fördelar som brukar förknippas med exempelvis autentiska prov och bedömningsprov när det gäller validitet, undervisning och lärande, måste man istället full ut acceptera "standardmodellen" och överge de "mätmodellens" krav på reliabilitet och normalfördelning av resultaten, hävdar Taylor (ibid.). Den svenska användningen av mera autentiska och öppna uppgiftsformat i de nationella proven, liksom anvisningarna i läroplanerna om en mera allsidig utvärdering av elevernas kunskaper för betygssättningen går inte i samma fälla som de amerikanska exempel som Taylor kritiserar, eftersom det svenska betygssystemet inte är norm- utan mål- och kunskapsrelaterat.

Pamela Moss (1994) tillhör dem som menar att det faktum att man inte kan tillämpa reliabilitetskriteriet vid alternativ bedömning inte alls nödvändigtvis utesluter, att sådana prov kan ligga till grund för viktiga beslut. Istället för kvalitetskriterier som hänför sig till psykometriska mätningar, menar Moss, att man bör använda sig av en hermeneutisk ansats när man vill försäkra sig om kvaliteten eller rimligheten i de slutsatser man drar av sina observationer av mänsklig handling. Det kan innebära att man istället för att sträva efter att olika bedömare ska uppnå samma resultat, låter dem mötas för att diskutera resultaten av sina respektive oberoende bedömningar tills de når enighet. Moss framhåller, att ett sådant förfaringssätt dessutom redan praxis i kritiska bedömningssituationer såsom domstolsbeslut och beslut om omhändertagande för sluten psykiatrisk vård. Samtidigt påpekar hon dock att metoden med bedömarkonferenser är omständlig, särskilt vid storskaliga bedömningar. Ett alternativ kan då vara att två personer utifrån vissa gemensamma kriterier gör oberoende bedömningar, som dokumenteras noga. Därefter kommenterar de varandras bedömningar utifrån sina tolkningar av tillgängliga data (ibid.). Exempel på liknande förfaringssätt från utbildningssammanhang är disputationer och uppsattsseminarier och från idrotten bedömning av simhopp och konståkning.

Faktorer som kan störa bedömningsresultatet

Bedömningskvaliteten, antingen den definieras utifrån traditionella kriterier för reliabilitet, validitet och generaliserbarhet, eller alternativa begrepp, är beroende av en mängd faktorer som är kopplade dels till provet som sådant och till tolkningen och värderingen av elevernas prestationer i förhållande till innehållet, dels till relationen mellan elev och prov. På båda dessa nivåer kan störningar uppträda som gör att resultaten blir påverkade, med vad som på engelska kallas "construct irrelevant variance" som följd. Med detta begrepp menas "andra systematiska skillnader i individuella egenskaper, än den kompetens som provet avser, vilka påverkar bedömningsresultatet" (Gipps & Murphy, 1994, s. 22). I detta kapitel tas några sådana potentiella "felkällor" upp.

Faktorer som påverkar prestationen

"Test-smartness"

Provresultaten kan bli missvisande om uppgifterna är sådana att eleverna kan lösa dem genom att genom att mobilisera andra kunskaper eller färdigheter än dem som provet är avsett att mäta. Cunningham (1998) påpekar emellertid, att eftersom det är svårt att veta exakt vilka kognitiva processer, som eleverna använder sig av när de besvarar en fråga eller löser en uppgift, så är det i praktiken ibland mycket svårt att konstruera proven på ett sådant sätt att de verkligen mäter just den kunskap, som man avser. Att en provuppgift är utformad för att mäta en viss kompetens innebär ju inga garantier för att den faktiskt gör det. Det finns exempelvis belägg för att elever, som har tränat på en viss uppgiftstyp använder sig av andra kognitiva strategier än de elever för vilka uppgiftstypen är ny (ibid. s. 133).

Birenbaum (1996) poängterar, att även komplexa och analytiska svar kan läras in och reproduceras. Madaus & O'Dwyer (1999) pekar på att elever kan lära in procedurer mekaniskt, så att de utan att någon egentligen förståelse finns, kan tillämpas korrekt i vissa välavgränsade sammanhang. Jones m fl (1999) och Stiggins (1999) rapporterar från amerikanska undersökningar, att det inte är ovanligt att lärare använder huvuddelen av undervisningstiden åt att förbereda eleverna på ett sådant sätt att de ska prestera bra på exempelvis nationella prov – d v s att klassens provresultat optimeras, inte de enskilda elevernas allsidiga kunskapsutveckling – blir det överordnade målet för undervisningen. Fokuseringen på proven kan förstås mot bakgrund av att inte bara prestige utan också ekonomiska resurser på flera håll i USA är knutna till provresultat. En konsekvens är naturligtvis att provens validitet försvagas.

Motivation och ängslan

Olika studier har visat att faktorer hos eleven, såsom inre motivation, grad av självständighet, inlärningsstil, och ”provängslan” har betydelse för i vilken utsträckning olika elever presterar på proven i enlighet med sin faktiska förmåga²⁹. Elevernas ängslan inför prov skulle kunna minskas genom att eleverna själva fick välja svårighetsgrad på provuppgifterna, menar Birenbaum (1996), och hänvisar till att detta i studier har visat sig vara fullt möjligt.

Eftersom man känner till att elevers möjligheter att kommunicera de kunskaper i en provsituation i hög grad beror på i vilken utsträckning som provuppgiftens art passar den enskilde, anser Birenbaum vidare, att man inom bedömningsforskningen bör ställa sig frågan om ”vilka personliga egenskaper, som påverkar

²⁹ Birenbaum (1996) refererar flera egna undersökningar om inlärningsstilens betydelse för inlärningsresultatet och sådana studier har också gjorts av bl a Säljö (1975), Marton (1986) och Marton & Booth (1997). Fransson (1977) och Pollard (1995) har visat hur nervositet och svagt intellektuellt självförtroende kan hämma elevers studieförmåga.

elevers preferenser för olika bedömningsformer och hur” (ibid., s. 20). Om elever får chansen att testas på de sätt som passar dem bäst, så kan man förvänta sig att det resulterar i att variansen till följd av ovidkommande faktorer minskar och möjligheterna att dra valida slutsatser om elevernas kunskaper ökar, menar Birenbaum (ibid.).

Biggs & Collis (1982) påpekar, att provprestationer även påverkas av faktorer, som är helt kopplade till individen och inte alls har med provet som sådant eller med bedömningsituationen att göra. Exempel på sådana faktorer är allmänna koncentrationssvårigheter och försvagad koncentration till följd av trötthet eller personliga bekymmer). Sådana ”felkällor” får naturligtvis störst genomslag vid bedömningar, som grundar sig på resultatet från ett enda provtillfälle. Om bedömningen av eleven sker på basis av data, som har insamlats vid flera tillfällen med hjälp av både olika typer av prov och informella observationer, så minskar risken att elevens kompetens underskattas av denna typ av skäl.

Språkliga hinder

Ett provresultat kan alltså påverkas av att provets utformning hindrar eleven från att fullt ut visa sina kunskaper. Exempelvis minskar essäprovs möjligheterna för elever med bristande skriftspråklig kompetens, att visa vad de kan. Om inte den skriftliga förmågan som sådan ska vara föremål för bedömningen, så kan kraven på att kunna uttrycka sig i skrift alltså ge upphov till ”mätfel” (construct irrelevant variance). Flervalssprov, som ju inte ställer sådana krav är på så sätt mer rättvisa mot dessa elever, men är ju begränsade till bedömning av konvergent tänkande. För att ge mindre skriftspråkskompetenta elever möjlighet att visa även divergent tänkande kan tillämpningsprov eller portfolio, d v s former som tillåter eleverna att demonstrera sina kunskaper på andra sätt än i text, vara ett alternativ. Även muntlig kommunikation ger naturligtvis ger naturligtvis bättre underlag för bedömning av mindre skriftspråkskompetenta elevers kunskaper. För elever med konstaterade läs- och skrivsvårigheter kan datorn vara användbar som ett hjälpmedel.

När det gäller elever från språkliga minoriteter kan bristande förtrogenhet med undervisningsspråket ibland hindra elever från att ge en fullödig bild av sina kunskaper. Objektiva svarsformat kan därför vara ett alternativ även för elever som inte har undervisningsspråket som första språk, liksom praktiska prov, och prov som bygger på andra uttrycksmedel än det språkliga. Ett tredje alternativ är att låta dessa elever formulera sig på sitt första språk, vilket naturligtvis förutsätter att den som bedömer själv förstår det aktuella språket eller har tillgång till en kompetent tolk.

Artificiella bedömningsituationer

Ett problem som är förknippat med "traditionella" prov är, att man måste "lita till elevernas förmåga att kunna "minnas och symboliskt representera kunskap (...) på begäran och i dekontextualiserade situationer" (Gordon & Bonilla Bowman, 1996, s. 33). Gordon & Bonilla Bowman menar, att man måste vara medveten om att det som krävs av eleverna i en typisk provsituation är något helt annorlunda än det tänkande, som eleverna är förmögna under optimala omständigheter och att vad eleverna kan klara av i en provsituation därför ofta inte säger så mycket om vad de skulle kunna klara i verkliga livet. Exempelvis Resnick (refererad i Gordon & Bonilla-Bowman, *ibid.*) har visat, att skillnaden mellan hur man uppfattar och kan handskas med kunskap inom och utanför skolan är betydande. Eftersom de erfarenheter, som eleverna har gjort utanför skolan har betydelse inte bara för hur de handskas med kunskap, utan också för hur de handskas med språket och förhåller sig till utbildning, så spelar också den sociokulturella bakgrunden roll för hur elever begriper provuppgifter av olika slag och vilka kompetenser de kan mobilisera för att lösa dem.

Författarna menar att *portfolio*, om det används medvetet och som en välintegrerad del av en konstruktivistisk undervisningskontext, kan antas ha en större potential än olika stickprovsbaserade bedömningsmetoder när det gäller att generera valid information om det unika i varje elevs kompetens och progression, eftersom denna modell gör det möjligt för eleven att

inkludera och dra slutsatser av kunskaper gjorda i olika sammanhang.

Broadfoot (1995) tar, som exempel på hur prov med starkt standardiserad form kan ge upphov osäkra resultat, upp försök som har gjorts i Storbritannien med att ge en sorts extern examination av samma typ, som sedan länge vid elva och tretton års ålder, till sjuåringar. Proven, som åtföljdes av instruktioner om att eleverna inte fick prata, samarbeta eller be läraren om hjälp, visade sig komma i konflikt med sociala, pedagogiska och etiska normer i de lägre stadierna. Barnen blev förvirrade och ibland sårade och lärarna kände sig degraderade. Detta motverkade den ursprungliga idén om att proven, genom att framstå som en helt vardaglig, integrerad del av klassrumsarbetet, skulle ge en så valid bild som möjligt av elevernas kunskaper. Då dessutom många lärare bröt mot reglerna och hjälpte sina elever, blev också reliabiliteten lidande, eftersom det fick som följd att eleverna kom att testas under olika villkor i olika klassrum. Resultaten blev därmed ojämförbara (Broadfoot, 1995).

Faktorer som påverkar bedömningsprocessen

Om bedömningen påverkas av stavfel och grammatiska fel, trots att stavning och grammatik inte har ingått som en del i de uttalade undervisningsmål, som provet är konstruerat för att mäta, försämras validiteten.

Birenbaum,(1996) varnar, när det gäller *portfolio* för risken att elever som inte är så duktiga på att t ex organisera stora material och göra snygga presentationer inte riktigt lyckas visa vad de kan, medan elever som besitter stora journalistiska färdigheter kan "blända" sina lärare. Framställningsteknik, självständighet och andra färdigheter, som man genom användningen av portfolio hoppas utveckla hos eleverna, kan alltså komma att betonas på ett sätt som gör att vissa elevers förståelse av ämnena kommer i skymundan.

Kulturella faktorer

Om resultaten från ett prov skiljer sig åt mellan olika samhällsgrupper och inte bara inom dem så kan det antingen bero på att provet eller bedömningen har givit en skev bild av kunskaperna hos dem som deltagit, eller på verkliga skillnader, som av en eller annan orsak faktiskt finns mellan elever som tillhör olika samhällsgrupper, ifråga om de kunskaper, som provet är avsett att mäta. Om skillnaderna i resultat mellan grupperna beror på provuppgifterna eller bedömningen snarare än på faktiska skillnader i elevernas kunskaper, så är de att betrakta, som "construct irrelevant variance", som beskrevs i det föregående avsnittet. När sådana störningar har effekter på gruppnivå dvs är systematiskt förknippade med en viss grupp, så betecknas de "test bias" (bias betyder ungefär "snedvridning") (Gipps & Murphy, 1994).

I psykometrisk forskning kan man mäta sådan "snedvridning" med hjälp av statistiska metoder, genom att kontrollera i vilken utsträckning olika grupper har klarat av enskilda uppgifter. Genom att byta ut vissa uppgifter, kan man eliminera gruppskillnader i utfallet (Gipps & Murphy, 1994). Exempelvis är IQ-testen konstruerade på ett sådant sätt, att män och kvinnor i medeltal får samma resultat.

Statistiska beräkningar av "bias" ger emellertid ingen information om *varför* resultatet skiljer sig mellan olika grupper. En förklaring till observerade skillnader i kunskapsstandard mellan olika grupper tar sin utgångspunkt i det faktum att existensbetingelserna skiljer sig åt för individer ur olika samhällsskikt och etniska grupper, och för flickor och pojkar, vilket innebär att eleverna utanför skolan lever delvis olika liv och gör olika erfarenheter. Dessa erfarenheter har bl a betydelse för hur de handskas med kunskap och med språket och för deras attityder till utbildning (t ex Bernstein, 1973, Jönsson m fl, 1994 och Berhanu, 2001). Därför spelar elevernas sociala och kulturella bakgrund en viktig roll för hur de begriper provuppgifter av olika slag och vilka kompetenser de kan mobilisera för att lösa dem. Att de sätt att kommunicera, de kulturella preferenser och

tillgångar och de fritidssysselsättningar, som är typiska för den västerländska medelklassen, ger de mest gynnsamma förutsättningarna för utveckling av sådana kunskaper som värdesätts i skolan, har genom åren konstaterats av många forskare. Tesfahuney (1999) förklarar fenomenet som en konsekvens av att de västerländska utbildningssystemen är "monokulturella". Med detta avser han att deras bildningsideal med rötter i upplysningen utgår från den vite, socialt privilegierade mannens perspektiv, vilket förutom att det ger just medlemmar av den vita, manliga medelklassen ett försprång i förhållande till andra grupper, innebär att exempelvis kvinnor och personer med utomeuropeiskt ursprung eller tillhöriga arbetarklassen definieras som avvikande och som bärare av en avvikande, mindre utvecklad kultur.

Även tillgången till utbildning skiljer sig åt mellan elever med olika kön och socioekonomisk, etnisk och även lingvistisk bakgrund. Exempelvis leder i storstäder, där den sociala och etniska bostadssegregationen är framträdande, till att elever med olika bakgrund får olika utbildning genom att de kommer att gå i skolor, där innehållet och de ekonomiska och kulturella resurserna skiljer sig åt på ett påtagligt sätt (Jönsson m fl, 1994). Att dessutom flickor och pojkar, barn till lågutbildade respektive högutbildade föräldrar, elever med svenska och utländsk bakgrund skiljer sig åt i fråga om val inom grundskolan och till gymnasiet³⁰ bidrar också till att den faktiska tillgången på utbildning skiljer sig åt mellan dessa grupper.

Gynnas eller missgynnas då elever ur socioekonomiskt svaga grupper och etniska och språkliga minoriteter av införandet av "alternativa prov"? Bilden som ges av olika empiriska studier ger inget entydigt svar på den frågan. Darling-Hammond & Aneesh (1996) pekar på studier som visar att amerikanska elever från skolor i områden med låg socioekonomisk status presterar sämre på "alternativa prov", både i förhållande till andra

³⁰ Detta visas i studier av, i tur och ordning, Staberg (1992), Erikson & Jonsson (1993) och Åhlund (1998).

grupper och i förhållande till ”traditionella prov”. De menar att en viktig orsak till detta är, att dessa elever inte har haft samma tillgång som andra till kognitiv och problemorienterad undervisning, eftersom de är överrepresenterade i specialundervisningsprogram, där man nästan uteslutande fokuserar på att förmedla ”baskunskaper” och använder flervalssprov.

För elever med annat modersmål, kan bristande förtrogenhet med undervisningsspråket utgöra ett hinder för att förstå innehållet i undervisningen, och därmed bidra till att de får ”mindre” utbildning än kamrater som har undervisningsspråket som modersmål (Gipps & Murphy, 1994). Att tvåspråkighet i sig skulle hindra elevers lärande, vilket det ibland har hävdats i debatten, finns det emellertid, enligt den kände språkforskaren Cummins (refererad i Gipps & Murphy, ibid.), inga belägg för. Däremot kan tvåspråkiga elevers möjligheter att skapa mening och förståelse hämmas i fall då man har underlåtit att utveckla deras första språk på ett fullödigt sätt. Om en elev saknar betydelsefulla begrepp i sitt första språk, som är det språk som eleven använder när hon eller han tänker, så orsakar detta kognitiv förvirring, menar Cummins.

Objektivitet och rättvisa i bedömningsprocessen

Det dominerande synsättet när det gäller rättvisa och objektivitet, skriver Gipps & Murphy (1994), utgår från att om alla testas under identiska förutsättningar och deras svar bedöms utifrån identiska kriterier, så kommer de resultatskillnader som uppstår att spegla de skillnader som finns mellan eleverna beträffande den kunskap, som provet mäter. För att inga ovidkommande faktorer ska påverka bedömningen bör eleverna enligt detta synsätt idealt vara anonyma för bedömaren, d v s bedömaren ska inte kunna koppla en viss prestation till en viss elev. Men även om bedömaren inte kan sammankoppla proven med individer, så finns det vid öppna svarsformat en risk att ”markörer” som språklig stil, handstil, vissa språkfel o. dyl. gör att bedömaren associerar svaret med en viss samhällsgrupp och omedvetet eller medvetet låter bedömningen påverkas av de egna värderingarna

om denna grupp, menar Gipps & Murphy (ibid.). Särskilt stor risk för felaktig bedömning löper elever med annan kulturell bakgrund än bedömaren. Detta problem är naturligtvis minimalt vid bedömning av prov med objektiva svarsformat, eftersom sådan bedömning sker mer eller mindre mekaniskt.

Det andra sättet att se på rättvisa i samband med prov är, innebär att den mest rättvisande bilden av elevernas kunskaper framträder när hänsyn tas till elevernas olikheter och varje elev får chansen att demonstrera sina kunskaper på det sätt, som passar honom eller henne bäst.

Att lärare, som själva tillhör majoritetskulturen, och som enbart bedömer elevernas arbete utifrån yttre kriterier, kan förbise kvaliteter i arbeten som gjorts av elever med annan etnisk bakgrund, illustrerar Moore (1993) med ett exempel där en lärare gör stora ingrepp i en bangladeshisk pojkes uppsats, som han ser som poänglös, orealistisk och ostrukturerad. Vad läraren inte förstod var att uppsatsen följde en *annan* struktur och beskrev en *annan* vardag än den som läraren hade erfarenhet av.

I det här fallet värderades alltså ett elevarbete oförtjänt lågt, inte i första hand för att eleven hade problem med undervisningsspråket eller med skrivandet, eller för att läraren var "orättvis", utan det viktigaste skälet för lärarens låga värdering var att han, tillhörande en annan kultur, inte hade de scheman som krävdes för att känna igen de kvaliteter som arbetet hade. Gordon & Bonilla-Bowman (1996, s. 49), skriver:

The value and the meaning of the writing does not exist somewhere on the page. It exists in the mind of the reader, which may be fertile and receptive to the message of the author or hostile and rejecting. The anticipation of the author's message often comes not from what is written on the page, but from shared experiences. One is more likely to have had shared experiences if both the reader and the author share a common culture.

Mot bakgrund av ovanstående, kan den självklara utgångspunkten att alla elever förtjänar att bli erkända för sina kunskaper fogas till raden av argument för att lärare med olika social och etnisk bakgrund arbetar i skolan.

Slutord

I tidigare kapitel har det framgått, att en stor del av de senaste par decenniernas bedömningsforskning har varit inriktad mot att undersöka hur traditionell bedömningspraktik påverkar undervisningen och elevers lärandestrategier och kunskapssyn och huruvida den är rättvis i förhållande till olika samhällsgrupper. Man har då exempelvis konstaterat, att de ”traditionella proven”, särskilt examinationer (d v s prov som görs i syfte att ge underlag för ackreditering eller urval och som har betydelse för den enskildes framtid, och där stor vikt fästs vid att resultaten ska vara jämförbara) medför en likriktning av undervisningen och en reproducerande inläring hos eleverna med ett ytligt och okritiskt förhållningssätt till kunskap som resultat, och desutom att de ofta missgynnar elever ur arbetarklassen och etniska minoriteter (ibland även det ena eller andra könet).

Man kan också notera en otillfredsställelse hos olika aktörer med det faktum att kognitiva och metakognitiva processer, divergent tänkande, social kompetens och liknande företeelser, som värderas högt i dagens västerländska utbildningsfilosofi och på arbetsmarknaden, inte kunnat mätas med de traditionella proven. Andra invändningar, som har riktats mot traditionella bedömningar handlar om, att de ger upphov till stress och till konkurrens mellan eleverna, och att de kan skada det intellektuella självförtroendet och motivationen hos elever som ofta misslyckas.

Med denna kritik mot de ”gamla” proven har, vilket vi också sett, ett intresse vuxit för alternativa former av bedömning, som hjälper läraren att individualisera och höja kvaliteten i sin undervisning, som stödjer elevernas utveckling av kognitiva, metakognitiva och sociala färdigheter, som är relevanta i förhållande till det aktuella läroplaner och till arbetsmarknadens nya kompetenskrav, som motiverar svagpresterande elever istället för att stämpla dem, och som är rättvisa i förhållande till olika samhällsgrupper.

Ideologiska perspektiv på ”det nya provparadigmet”

En väsentlig del av upphovet till de traditionella provens negativa konsekvenser har tillskrivits de reliabilitetskrav som är förknippade med summativa bedömningssyften; kraven på reliabla, jämförbara resultat har lett till bedömningsituationer som har gjort eleverna nervösa, till uppgifter med låg innehållslig relevans, etc. Som en röd tråd genom denna argumentation har viljan att ersätta summativa bedömningar med formativa skymsat, liksom antagandet att om man bytte ut de prov, som är till för att kontrollera, certifiera, ackreditera, sanktionera, sortera o s v mot prov som är till för att hjälpa, utveckla och uppmuntra, så skulle man få en utbildning, som inte bara mera är mer human och rättvis, utan också mer kvalificerad och effektiv.

Det är emellertid viktigt att uppmärksamma att utvecklingen av nya, autentiska och mera kvalitativt inriktade prov, och intresset för kognitiva och metakognitiva teorier om lärare har drivits utifrån skilda ideologiska utgångspunkter – men detsamma gäller också kritiken mot alternativa prov och försvaret av traditionella!

Konservativa argument för nya former av prov kan man bl a finna hos George Cunningham. I sin bok *Assesment in the Classroom – Constructing and Interpreting tests* (1998), som är avsedd att användas som kurslitteratur för blivande lärare motiverar han behovet av nya och mera varierade bedömningssätt med, att den stora kriminalitet, det växande drogmissbruk och de många oönskade graviditeter och könssjukdomar som konstaterats bland skolungdomar ibland har skyllts på att skolan inte är bra. Han hänvisar också till att flera statliga rapporter (mest uppmärksammas blev den under Reagans regering tillkomna ”A Nation at Risk”) pekat på att kunskapsstandarden bland amerikanska ungdomar i förhållande till ungdomar från andra länder är anmärkningsvärt låg. De som har uttryckt sin oro över de moraliska och kunskapsmässiga bristerna hos ungdomen har, enligt Cunningham, ofta efterlyst mera och eller

annorlunda prov i skolan som ett medel för att ta kontroll över elevernas lärande, sätta in åtgärder där det behövs på ett tidigt stadium och motivera eleverna³¹.

Då det är i linje med en konservativ utbildningssyn att prioritera överförandet av ett dominerande ”kulturarv” till de yngre generationerna och att upprätthålla traditionella ämnesgränser (se Englund, 1996), är det emellertid även logiskt att utifrån ett sådant perspektiv förespråka traditionella prov med inriktning mot konvergent tänkande och återgivning, framför sådana som är tvärdisciplinära och inriktade på att ”skapa ny kunskap”.

Utvecklingen av alternativa prov har även drivits med marknadsliberala motiv. Drivkraften för att förändra utbildningen och bedömningsformerna har då haft sitt upphov i förändringar, som skett på arbetsmarknaden, främst på tekniska innovationer inom industrin, och som lett till nya krav på presumtiva arbetstagares kompetens och egenskaper (Gordon & Bonilla-Bowman, 1996). Ur detta perspektiv har de standardiserade provens validitet kritiserats med hänvisning till att de inte i första hand mäter sådana kunskaper, som ”näringslivet” har behov av. Vad som är värdefull kunskap definieras i detta perspektiv alltså av ”näringslivet” (ibid.).

Olika ”alternativa” former av prov har också på senare år fått genomslag i storskaliga utvärderingar, exempelvis på nationell nivå. I både Storbritannien och USA, där, som har nämnts tidigare, resultaten från sådana utvärderingar ofta är knutna till

³¹ Både Hargreaves (1989) och Frykman (1998) drar paralleller mellan dagens sublima försök att kontrollera elevernas tankar och känslor och Foucaults skrift om övervakning av fångar i ett fängelse ”panopticon”, där eleverna när som helst riskerar att få väktarens ögon på sig, men genom att övervakningen sker genom ett envägsfönster, så vet de aldrig när de faktiskt har det. Det finns dessutom inget sätt att undkomma övervakningen. Det innebär att eleverna alltid måste uppträda som om de vore iakttagna. I den gamla skolan kunde eleverna skydda sig mot kontrollen genom att gömma sig i sig själva, medan de lät sin yttre gestalt, kroppen, underkasta sig den ordning som lära- ren bestämde.

ekonomiska resurser, har det visat sig att proven, trots sin nya utformning och trots förväntningar om motsatsen, har haft en utarmande effekt både på innehåll och metoder (t ex Madaus & O'Dwyer, 1999). Idén har varit att skolor och distrikt (ibland även lärare) genom att på basis av provresultat konkurrera om resurser, ska bli "effektiva". Det är för detta ändamål som t ex den tidigare beskrivna "Kentucky-modellen" utarbetats (Broadfoot, 1996).

Också i den politiska vänstern, där utjämningen av sociala skillnader antas vara det överordnade målet, är inställningen till "det nya provparadigmet" kluven. Det perspektiv, som varit vägledande i traditionell socialdemokratisk utbildningspolitik, bygger på en kvantitativ syn på jämlikhet, där det viktigaste målet varit att eliminera könstillhörighetens och den sociala bakgrundens betydelse för individers möjligheter att nå olika sociala positioner (Åberg, 1992). Enligt Gordon & Bonilla-Bowman (1996) fokuseras i detta perspektiv eleven emellertid som individ, "utan både klassmässig och etnisk kontext":

De tror att i ett system, som erbjuder alla elever möjligheten att lära sig, kommer de mest kapabla, att stiga till toppen, oavsett sitt ursprung. En utbildning som utvecklar sådana intellektuella färdigheter som värdesätts av de dominerande i samhället, kommer att göra eleverna förmögna att lyckas på utbildningsarenan och på så sätt utrusta dem för att konkurrera framgångsrikt i den ekonomiska sfären.

Gordon & Bonilla-Bowman (1996, s. 34)

I detta perspektiv är det främst i egenskap av sorteringsinstrument, som bedömning och undervisning diskuteras ur jämlikhetssynpunkt. Man kan säga att resonemanget bygger på en hegemonisk kunskapssyn, där innehållets relation till olika samhällsgrupper sällan analyseras och samhällets sociala struktur i princip tas för given.

Ur ett sådant perspektiv på jämlikhet är det eftersträvansvärt med en sådan pedagogik och sådana bedömningsformer, som underlättar för elever med "icke-akademisk bakgrund", att lyckas i utbildningssystemet *så som det ser ut*. Med den utgångspunkten har skolforskare och -debattörer, ofta med hänvisning till

Bernsteins teorier (t ex 1973³²), argumenterat för en undervisning, som går från det enkla till det sammansatta och som utgår från väl definierade mål och procedurer. Dessa berör bl a hur sättet att använda språket i olika samhällsskikt stämmer överens med olika typer av undervisning. Progressiva undervisningsansatser har ur detta perspektiv kritiserats för att utgå från en ”dold läroplan”, som menas bygga på medelklasstyypiska förhållningssätt till utbildning, kunskap och språk, och därmed missgynnar elever ur ”arbetarklassen”:

Barn utanför ”den dominerande kulturen” kommer inte till skolan med en internaliserad kod, som svarar mot vad de förväntas göra i skolan, så som många barn från medelklassen gör. Av den anledningen kan det vara så att explicit undervisning, i förhållande till processororienterade ansatser, bättre tjänar dessa elevers behov. (Gordon & Bonilla-Bowman, 1996, s. 48)

Delpit (refererad i Gordon & Bonilla-Bowman, *ibid.*) har uppmärksammat att föräldrar ur mindre privilegierade samhällsgrupper och som tillhör språkliga och etniska minoriteter ofta är oroliga för att skolan inte ska utrusta deras barn med de kun-

³² 42 Samlingskoden bygger på en atomistisk kunskapssyn och en inställning att man måste lära sig det enkla före det sammansatta, att man kan närma sig en förståelse av innehållets djupstruktur bara genom att ”beta av” lager efter lager från ytnivån. I utbildningssystem som är starkt präglade av en samlingskod kommer flertalet aldrig att nå de nivåer i innehållet där principerna för skapandet av ny kunskap (eller, med Bernsteins ord. ”nya verkligheter”) uppenbaras. Istället tvingas de till mer eller mindre reproducerande inläring av kunskaper som andra tillhandahåller. Det är först mycket sent i utbildningskarriären, som man har kvalificerat sig för att handskas med kunskap på djupstrukturnivå, och det är bara några få som överhuvudtaget når dithän. Undervisning enligt den integrerade koden är däremot från början riktad mot innehållets djupstruktur. Olika undervisningsinnehåll är inte skilda åt genom mekaniska regler eller traditionella ämnesuppdelningar. Inriktningen mot djupstruktur innebär istället att man söker efter efter mönster och analogier som kan förklara företeelser och sammanhang som uppfattats på ytligare nivåer. Därför ses det som lämpligare att organisera innehållet utifrån sådana tematiseringar och avgränsningar, som gör det möjligt att se sådana mönster, än utifrån traditionellt disciplinära gränser. Med kunskapssynen i samlingskoden följer relativt entydiga kriterier för kunskapsutvärdering. Det viktigaste skälet för detta är den tydliga åtskillnaden mellan elevers och lärares sk vardagskunskap eller personliga erfarenheter och skolkunskap; bara kunskaper som är oberoende av all personlig kontext rymts i undervisningen.

skaper som krävs i samhället. Delpit talar om "gate-keeping points", olika tillfällen på vägen genom utbildningssystemet och i arbetslivet, då individer bedöms och sorteras inför framtiden (exempelvis examina, antagningsprov och anställningsintervjuer); att inte låtsas om att sådana tillfällen existerar och är betydelsefulla för individen vore, menar Delpit, att göra dessa grupper orätt. Exempelvis menar hon, att medan stavning kanske inte ses som ett särskilt viktigt kriterium för medlemmar av den dominerande kulturen, så kan det för personer från andra kulturer utgöra en del av "inträdesbiljetten" (ibid.).

Ett relativt sent tillskott till kritiken mot progressiv pedagogik när det gäller jämlikhet, utgör Frykmans bok "Ljusnande Framtid" (1998). Frykman skriver där att den moraliska panik över skolan (gällande mobbning, respektlöshet, lärarnas utbrändhet etc.), som beskrivs i och underblåses av media, skymmer en analys av skolans problem som tar sin utgångspunkt inte i faktorer utanför skolan och utanför skolans kontroll (att barnen mår dåligt för att föräldrarna är arbetslösa, att klasserna är så heterogena, att tidsandan är pessimistisk, att föräldrarna inte uppfostrar sina barn, att man har gjort så stora ekonomiska nedskärningar osv), utan *i skolan*.

Frykman menar att skolans syfte har förändrats – från att ha varit en plats dit eleverna kom för att tillägna sig kunskaper som de inte kunde få någon annanstans, med siktet inställt på att kunna använda dessa kunskaper och färdigheter i framtiden för att skaffa sig ett bra yrke, så har skolan gjorts till en plats för "identitetsarbete", där eleverna ska lära känna sig själva och utveckla vissa värderingar och personlighetsdrag. Man är i skolan, skriver Frykman, "för att bli någon" och inte, som förr, "något". Den gamla, starkt ritualiserade skolan tillät och krävde av eleverna (och lärde dem dessutom) att visa sig enbart i egenkap av elever, att spela en elevroll och utifrån denna kunde barn med olika bakgrund delta i "spelet" på lika villkor.

Frykman tar också upp de förändringar i synen på kunskap som har skett i övergången från den gamla, ritualiserade skolan till den nya, "personlighetsutvecklande", och som innebär att kun-

skapen alltmer kommit att förknippas med ”den sociala personen”, istället för att betraktas som fristående. Detta sker, menar han, genom att ”lärlingen” (eleven) deltar i den praktik som hon eller han ska lära sig, och innebär att kunskapen tillägnas oreflektat och okritiskt – den blir en del av elevens identitet³³.

De invändningar som ur jämlikhetssynpunkt har riktats mot progressiv undervisning kan naturligtvis även överföras till resonemang om bedömning. Ur det perspektiv på jämlikhet där man ser utjämning av chanser mellan individer ur olika grupper när det gäller att konkurrera om betyg och sociala positioner, kan därför konventionella bedömningsmetoder ibland synas vara att föredra framför mera progressiva.

Ur ett mera radikalt jämlikhetsperspektiv är ambitionen vidare än att likställa chanserna för individer att nå olika positioner i den sociala samhällshierarkin – den syftar till att utmana de själva de tankeformer, som legitimerar och bidrar till att upprätthålla den sociala skiktningen i samhället. Vare sig kunskapsinnehållet eller samhällsordningen ses alltså som konstanta i detta perspektiv, som fokuserar varje individs rätt till en nyttig och meningsfull utbildning på sina egna kulturella villkor, en utbildning, som inte definierar världen ur de dominerande gruppernas perspektiv, utan bygger på en medvetenhet om kunskapens historiska och politiska kontext och som respekterar alla elevers erfarenheter och kunskaper.

Centralt för denna utbildningssyn är också avståndstagandet från teorier om intelligens som en medfödd och konstant egenkap hos individen. Istället utgår man från att intelligensen är utvecklingsbar. De processer som är involverade i komplext tänkande ska göras synliga och därigenom möjliga att tillägna sig

³³ Här kan man jämföra med Kvaless resonemang (1992) om ”tekniska”, ”auktoritära” och ”terapeutiska” examinationsformer. Kvale menar att ingen av dessa former möjliggör diskussioner om kunskapens validitet – den ”terapeutiska” av den anledningen att den bygger på tesen att varje sanningsanspråk är grundat i individernas subjektiva världsuppfattning och därför inte kan diskuteras och värderas på saklig grund.

för alla (Wolf & Reardon, 1996). Vidare menar Wolf & Reardon (ibid., s. 7), att proven inte kan diskuteras skilt från undervisningen:

Innan man kan förvänta sig nya reliabla bedömningsformer, så måste vi öppna diskussionen om vad som gör ett bra arbete bra; göra det möjligt för lärare att utveckla gemensamma språkbruk för att beskriva de kvaliteter och problem, som de ser i olika elevarbeten och integrera bedömningen i undervisningen så att eleverna förstår poängen med sitt kontinuerliga arbete med att förbättra sina arbeten och med att visa att de förstår.

Prov i praktiken – konsekvenser

Förespråkare för denna mera radikala och innehållsrelaterade kritik av utbildningssystemet (t ex Gordon & Bonilla-Bowman; Darlington & Aness, Broadfoot) framhåller ibland, att dessa utbildningsfilosofiska utgångspunkter är oförenliga med de aspekter av undervisning och kunskapsbedömning, som är kopplade till sortering av elever. Inställningen till hur man konkret ska hantera konsekvenserna av dessa aspekter av skolans samhälleliga roll i ett stratifierat samhälle, skiljer sig dock mellan de författare som delar denna utgångspunkt.

En slutsats är, att man i den dagliga undervisningen bör dra nytta av de alternativa provens pedagogiska fördelar, medan man när det gäller exempelvis examinationer, antagnings- och urvalsprov bör fortsätta förlita sig på konventionella prov, eftersom man av rättssäkerhetsskäl inte anser sig kunna undvara dessas överlägsna psykometriska kvaliteter (t ex Broadfoot, 1995; Gipps & Murphy, 1994).

En annan slutsats är, att skolan bör avstå från sådana prov där mycket står på spel för individen och där exakthet och jämförbarhet i resultaten är av avgörande betydelse. Wolf & Reardon (1996, s. 28) menar, att vi först av allt borde

(...) koncentrera oss på att utveckla gemensamma språk för att kunna beskriva elevers kunskaper och hur de använder det i kreativa, vardagliga sammanhang som är relevanta ur undervisningssynpunkt, och som erkänner elevers individualitet, lärares autonomi, komplexiteten i de handlingar

där eleverna uttrycker sin kunskap, och de sociokulturella sammanhang vari proven ska begripas.

Vilka konsekvenser skulle då dessa båda vägar kunna få i praktiken? När det gäller det första, så visar empiriska studier, att i de fall då undervisningen och proven har olika inriktning, så är det provens inriktning som har störst betydelse för vad eleverna lär sig och vilken kunskapssyn de utvecklar (Marton, 1986; Unenge, 1994; Beach, 1999). Att eleverna i första hand riktar sina ansträngningar mot att förvärva sådana kunskaper, som de antar värderas vid betygssättning eller examinationer, måste ses som naturligt med tanke på den betydelse som betygen har för deras framtid; ju större ansträngningar som krävs av eleverna för att nå de högsta betygen, desto mindre kan man anta, att de prioriterar att lära sig sådant, som de inte betraktar som relevant för betygen. Om undervisningsmålen och den formativa kunskapsbedömningen skulle fokusera andra kunskaper och ha andra teoretiska utgångspunkter än de prov, som syftar till ackreditering och sortering, skulle proven heller inte ge underlag för valida slutsatser om elevernas kompetens i förhållande till undervisningsmålen.

När det gäller det andra alternativet – att avskaffa betygsgrundande bedömning och därmed befrias från de reliabilitetskrav, som de är förknippade med och som har oönskade konsekvenser för elevernas lärande, så riskerar det att leda till den paradoxala situationen, att elevernas lärande får en än snävare inriktning: Om skolan fränsäger sig uppdraget, att examinera elevernas kompetens i förhållande till de undervisningsmål, som definieras av politiskt beslutade läroplaner, kan det leda till att den makt som staten har haft, att definiera innehållet i utbildningen i praktiken går förlorad. Om man antar, att en majoritet av eleverna intresserade av att investera sin tid och sitt arbete på det sätt som de tror ger bäst social och ekonomisk utdelning i framtiden, så kommer deras föreställningar om vad som är gångbar kunskap på arbetsmarknaden (eller den utbildning, som de är intresserade av att kvalificera sig för), att vara mera vägledande för vad de lär sig än den formella läroplanen. En följd av detta skulle kunna vara, att eleverna i praktiken får en snävare utbild-

ning än vad som annars hade varit fallet (för elever med siktet inställt på tekniska yrken, vore det t ex "oekonomiskt" att ägna tid åt att förkovra sig i exempelvis historia). Makten över innehållet i skolan skulle därmed förskjutas från den politiska till den ekonomiska sfären.

I vilken mån återspeglas då det paradigmskifte, som man talar om i litteraturen om bedömning, i den pedagogiska vardagen? Det är svårt att säga – troligen ser det mycket olika ut på olika skolor i Sverige. Den litteratur som ingår i detta arbete ger inte underlag för att besvara frågan. Att de alternativa bedömningsformer och nya sätt att se på bedömning åtminstone inte hunnit bli norm i amerikanska klassrum vid en tidpunkt då forskningen om dem befann sig på en höjdpunkt, indikeras av Cizeks uttalande (1997, s. 4), om att "den *behaviouristiska hegemonin* i hög grad lever kvar i praktiken", liksom Gordons & Bonilla-Bowmans (1996) påpekande, att kriterierelaterade, tillämpningsbaserade bedömningsmodeller trots att de har existerat i cirka trettio år, ännu inte i nämnvärd utsträckning har påverkat "testindustrins" produkter i fråga om idéer och tekniker.

Gordon & Bonilla-Bowman menar, att de moderna proven fortsätter att fokusera resultat på bekostnad av processer och att prioritera sortering, förutsägelse och urval framför att stödja undervisning och lärande. Också Shepard (2000) påpekar att de i mätforskning och behaviourism grundade synsätten på bedömning lever kvar i klassrummen, trots att undervisningen på senare år i ökande grad har införlivat konstruktivistiska och kognitiva synsätt, och pekar som en förklaring på det faktum att lärares, föräldrars och elevers föreställningar om undervisning och bedömning har en mycket stark förankring i vår kultur och därför är svåra att ändra. "Varje försök att ändra formen och syftet med bedömningen i klassrummet för att på ett grundläggande sätt göra den till en del i inlärningsprocessen måste erkänna kraften av seglivade och dolda föreställningar", skriver Shepard (ibid., s. 9). Hon ser dock att förändringen är möjlig.

I Sverige har den standardiserade, storskaliga testningen av elever inte varit lika utbredd som framförallt i USA. Här har, frånsett ett kort mellanspel med undervisningsteknologi där det var

läromedelsförlagen som kontrollerade innehållet och bedömningen av elevernas kunskaper, alltid haft relativt stor autonomi på det lokala planet för lärarna att utforma prov, och det har, sedan real- och studentexamen avskaffades, helt och hållet varit lärarnas bedömningar som legat till grund för betygen (även om de nationella proven och, tidigare, standardproven förvisso har haft ett stort inflytande över lärarnas beslut om betyg). Att eleverna själva formulerar sina svar på provfrågor är heller inget nytt, utan har förmodligen alltid varit det dominerande svarssättet framför t ex flervalsprov.

Den svenska erfarenheten visar emellertid att det faktum att lärarna själva konstruerar proven och att eleverna själva konstruerar svaren, inte på något sätt är oförenliga med den traditionella reciterande pedagogiken, med behavioristisk inlärnings-teori, positivistisk epistemologi eller en atomistisk kunskaps-syn – allt det som förmenta anhängare av ”det nya provparadigmet” vill göra upp med. Inte heller utgör dessa egenskaper tillräckliga förutsättningar för en konstruktivistisk kunskaps-syn hos vare sig elever eller lärare. Det finns inget automatiskt samband mellan lokalt utformade prov och laborativa, utmanande undervisningsformer, ”djupinläring”, metakognition eller kritisk reflexion. Inte ens elevernas ”forskning” kring fritt valda eller tematiska ämnen, ett vanligt förekommande inslag i den svenska skolan särskilt i de lägre åren, leder nödvändigtvis till att sådana utbildningskvaliteter utvecklas, utan resulterar ofta i återgivande och tillägnande av lösryckta fakta. Detsamma gäller grupparbeten och muntliga redovisningar³⁴. Sambandet mellan form och innehåll är aldrig givet.

Formella system för kunskapsbedömning har alltså spelat en viktig roll för att öka den sociala mobiliteten i moderna samhällen och låta ”det sociala arvet” ersättas av mindre godtyckliga faktorer när det gäller människors chanser att som vuxna komma i åtnjutande av ekonomiska, sociala och kulturella tillgångar. Samtidigt är det viktigt att inse, att proven i praktiken faktiskt ofta har missgynnat elever ur traditionellt underprivili-

³⁴ För exempel och diskussion, se Eriksson & Korp (1996)

gierade samhällsgrupper och därmed bidragit till att legitimera skolans roll i den sociala reproduktionen. Samhällen med ett utbildningssystem, som bygger på teorier om att de utbildningsmeriter som är avsedda att ligga till grund för senare fördelning av välstånd av olika slag, till väsentlig del är beroende av människors medfödda intelligens, kan dessutom knappast betraktas som demokratiska. Vissa utbildningsforskare har, som tidigare nämnts, sett det som en demokratisk nödvändighet, att teorier om inneboende och konstant intelligens överges som bas för utbildnings- och bedömningsystem, och att provens roll att stödja pedagogiska processer prioriteras framför sorterande funktioner. Först då, har man resonerat, kan lärare få chansen att skapa optimala villkor för lärande för elever med olika inlärningsstilar och preferenser och ur olika samhällsgrupper. I Sverige öppnades möjligheten för detta i och med övergången till ett mål- och kriterierelaterat betygssystem. Ändå har de sociala skillnaderna i betyg inte minskat, under de första åren (Skolverket, 2001).

Frågor för fortsatt forskning

Forskningen i sig kan inte ge svar på vilka frågor som framtida forskning bör ägnas åt att undersöka. Däremot kan "kunskapsluckor" bli synliga i en genomgång som denna av forskningsresultat och vetenskapliga diskussioner inom ett område. "Kunskapsluckorna" i svensk bedömningsforskning är dock endast en liten delmängd av "det vi ännu inte vet" om kunskapsbedömning. Inom alla områden kan den kognitiva överbyggnad som vetenskapen utgör liknas vid toppen av ett isberg, som i övrigt består av andra former av vetande, t ex "vardagsvetande" och "kunskap i handling", kunskaper som existerar i form av kulturella normer, språkliga diskurser, institutionella praktiker, "smaker" och livsstilar och i människors kroppar. Det är först när den befintliga forskningen inte befinns tillräcklig eller av rätt sort för att stödja önskade förändringar, som "kunskapsluckorna" i forskningen uppstår. Frågor om vilka förändringar som är önskvärda måste till slut återföras på värderingar om det goda livet, det goda samhället och, när det gäller bedömning av elever, den goda skolan och det meningsfulla lärandet.

Den vetenskapliga överbyggnaden om bedömning i svensk utbildning är över huvudtaget liten, även om man naturligtvis kan dra många generella slutsatser av den internationella forskningen. Men i vilken riktning bör den byggas ut?

Ur perspektivet att det är önskvärt att professionella grupper som på olika sätt är engagerade i skolan – det må vara lärare, politiker, lärarutbildare, forskare, rektorer eller läromedelsproducenter – arbetar för förändringar som innebär att elever med olika social och kulturell bakgrund och kön får samma chans till utbildning av hög kvalitet, framstår det som viktigt att dels definiera vad ”hög kvalitet” innebär för olika grupper, dels att studera grupper av elever som kommer till korta i utbildningssystemet och vad i samspelet mellan inläring, bedömning och undervisning som är av kritisk betydelse för att ”lyckas”. Några utgångspunkter för att gå vidare med denna slags forskning ges i Skolverkets rapport (2001) *Utan fullständiga betyg – varför når inte alla målen?*

Innehåll och former i såväl undervisning som bedömning behöver också problematiseras i förhållande till olika samhällsgruppers legitima strävanden efter ett gott liv. Här behöver läroplans-, kunskaps sociologiska och didaktiska teorier kombineras i empiriska studier av undervisning och bedömning i olika sociala miljöer. Vi behöver få kunskap både om ”läget” i stort för att få en bild av hur sociala strukturer i samhället medieras, reproduceras och utmanas av elever och lärare i bedömnings situationer i klassrummet och om ”goda exempel”, som kan ge en bild av potentialen i förändrade bedömningspraktiker.

Vi behöver också ökade kunskaper om

- hur frågor om bedömning hanteras inom ramen för lärarutbildning i Sverige
- hur kunskapsbedömning i t ex praktisk-estetiska ämnen och inom ramen för yrkestekniska ämnen i gymnasiet sker och om bedömningsmodeller från dessa ämnen kan bidra till utvecklingen av mera autentiska bedömningsstrategier strategier inom andra ämnen

- relationen mellan de olika ämnenas kunskapsstruktur och traditioner å ena sidan och bedömningspraktiker å den andra
- vad en ökad användning av informella observationer, själv- och kamratbedömning som grund för betygssättning innebär för kommunikations- och maktförhållanden i klassrummet
- hur bedömningen i klassrummet påverkar elevernas inlärningsstrategier, motivation och deras konstruktion av sig själva som lärande, samt deras attityder till utbildning
- hur formativ bedömning kan användas som ett verktyg för elever från skilda språkliga och kulturella grupper för att utveckla sin kompetens att tolka och hantera såväl skolans mål och majoritetskulturens normer, som egna och andra kulturella sammanhang
- vilka kvaliteter i elevernas lärande som lärarna fokuserar i bedömningen (både formativ och betyggrundande), vilken typ av data som utgör underlag för lärares betygssättning och hur de tolkar dessa data i relation till betygskriterierna

Referenser

BALL, S.J. & LARSSON, S. (Red.)(1989), *The struggle for democratic education and participation in Sweden*, New York, Philadelphia, London: The Falmer Press.

BANKS, J. (1999). *An introduction to multicultural education*. Boston: Allyn & Bacon.

BEACH, D. (1999). Om demokrati, reproduktion och förnyelse i dagens gymnasieskola i *Pedagogisk forskning i Sverige*. Årg. 4, Nr 4 (s. 349 – 366).

BEACH, D. (1995). *Making Sense of the Problems of Change*. Göteborg: Acta Gothoburgensis.

BENGTSSON, J. (1996). Vad är reflektion? Om reflektion i läraryrke och lärarutbildning. I C. Brusling & G. Strömqvist (red.) *Reflektion och praktik i läraryrket*. Lund: Studenlitteratur.

BERHANU, G. (2001). *Learning-In-Context. An Ethnographic Investigation of Mediated Learning Experiences among Ethiopian Jews in Israel*. Göteborg: Acta Gothoburgensis.

BERNSTEIN, B. (1973). *Class, Code and Control, Volume 3: Towards a theory of Educational Transmissions*. London, Henley och Boston: Routledge & Kegan Paul.

BIGGS, J.B. & COLLIS, K.F. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy* New York; Academic Press.

BIRENBAUM, M. (1996). Assessment 2000: Towards a pluralistic Approach to Assessment. I M. Bierenbaum & F. Dochy (red.) *Alternatives in Assessment of Achievement, Learning processes and Prior Knowledge*. Massachusetts: Kluwer Publishers.

BLACK, P. , WILLIAM, D. (1998). *Inside the Black Box*. London: Kings University.

BLACK, P. (1999). Assessment; Learning Theories and Testing Systems. I *Learners, Learning and Assessment* (p. 118-134). London, Thousand Oaks och New Dehli: Sage Publications.

BLOOM, B. S. M. FL. (Red.) (1956). Taxonomy of Educational Objectives i *Handbook 1. Cognitive Domain*. New York: McKay.

BONNIOL, J. J. (1991). The Mechanisms Regulating the Learning Process of Pupils – Contribution to a Theory of Formative Assessment. I P. Weston (ed.) *Assessment of Pupil Achievement: Motivation and School Success*. Rapport från the Educational research workshop som hölls i LiËge 1989. Amsterdarn/Lisse: Swets & Zeitlinger B. V.

BROADFOOT, P. (1995). Performance assessment in perspective: international trends and current English experience. I Torrance (red.) *Evaluating Authentic Assessment*. Buckingham, Philadelphia: Open University Press.

BROADFOOT, P. (1996). *Education, assessment and society – a sociological analysis*. Buckingham: Open University Press.

BROADY, D. (1982). *Den Dolda Läroplanen*. Stockholm: Symposion.

BROADY, D. (1994). Bildningsfrågan. I *Bildning och kunskap*. Särtryck ur läroplanskommittËns huvudbetänkande SOU 1992:94 Skola för bildning. Stockholm: Liber.

CALFEE, R.C., MASUDA, W. V. (1997). Classroom Assessment as Inquiry. I G. D. Phye (red.) *Handbook of Classroom Assessment. Learning, Adjustment and Achievement*. San Diego, London, Boston, New York, Sydney, Tokyo och Toronto: Academic Press.

CARLGREN, I. (1994). Kunskap och lärande. I *Bildning och kunskap*. Särtryck ur läroplanskommittËns huvudbetänkande SOU 1992:94. Skola för bildning. Stockholm: Liber.

- CHOPPIN, B.H. (1990). Evaluation, Assessment and Measurement. I H. Walberg & G. Haertel (red.). *The International Encyclopedia of Educational Evaluation*. Oxford: Pergamon Press.
- CIZEK, G.J. (1997). Learning, Achievement and Assessment. I G.D. Phye (red.) *Handbook of Classroom Assessment. Learning, Adjustment and Achievement*. San Diego, London, Boston, New York, Sydney, Tokyo och Toronto: Academic Press.
- CLAESSON, S. (1999). *Hur tänker du då? Empiriska studier om relationen mellan forskning om elevuppfattningar och lärarens undervisning*. Göteborg: Acta Universitatis Gothoburgensis.
- CUNNINGHAM, G. (1998). *Assessment in the Classroom, Constructing and Interpreting Tests*. London och Washington D. C.: The Falmer Press.
- DAHL, E-L. (1999) *Kyrkan, staten och reformpedagogiken*, manus.
- DARLING-HAMMOND, L. & ANCESS, J. (1996). Assessment and School Development. I B. Boykoff & D.P. Wolf (red.), *Performance-based Student Assessment: Challenges and possibilities*. Ninety-fifth Yearbook of the National Society for the study of Education. Chicago: The University of Chicago Press.
- DONALDSON, M (1979). *Hur barn tänker*. Lund: Liber Läromedel.
- ECKSTEIN, M.A. & NOAH, H.J. (1993). *Secondary School Examinations*. New Haven och London: Yale University Press.
- EISNER, E. (1985). *The Art of Educational Evaluation: A Personal View*. London och Philadelphia; The Falmer Press.
- EISNER, E. (1991). Taking a Second Look: Educational Connoisseurship Revisited. I M. McLaughlin & D. Phillips (red.), *Evaluation and education at quarter century* (s. 169-187). Chicago: The National Society for the Study of Education.

ELLIOTT, J. (1998). *The Curriculum Experiment: Meeting the Challenge of Social Change*. Buckingham/Philadelphia: Open University Press.

ENGLUND, T. (1996). *Utbildningspolitiskt systemskifte?* Stockholm: HLS.

ERIKSSON, G. (1998). Test-taker feedback – elevers synpunkter som del av arbetet med nationella prov i engelska. I B. Ljung & A. Pettersson (red.) *Perspektiv på bedömning av kunskap*. Stockholm: Lärarhögskolan i Stockholm, Institutionen för pedagogik.

ERIKSSON, L. & KORP, H. (1996). En hund i backspegeln. I C. Brusling & G. Strömqvist (red.), *Reflektion och praktik i läraryrket*. Lund: Studentlitteratur.

ERIKSON, R. & JONSSON, J. (1993). *Ursprung och utbildning*. SOU 1993: 85. Stockholm: Fritzes förlag.

ERIKSON, R. (1992). Lika möjligheter? Livschanser och social bakgrund. I R. Åberg (red.), *Social bakgrund, Utbildning, Livschanser*. Stockholm: Carlssons Bokförlag.

FRANSSON, A. (1978). *Att rädas prov och vilja veta*, Göteborg: Acta Gothoburgensis.

FRYKMAN, J. (1998). *Ljusnande framtid*. Lund: Historiska media.

GIPPS, C. & MURPHY, P. (1994). *A Fair Test? Assessment, achievement and equity* Buckingham/Philadelphia: Open University Press.

GIPPS, C. (1994). *Beyond Testing – Towards a Theory of Educational Assessment*. London: The Falmer Press.

GIPPS, C. (1997). *Assessment: a teacher's guide to the issues*. London: Hodder & Stoughton.

- GLASER & SILVER (1994). Assessment, Testing and Instruction: Retrospect and prospect. I *Review of research in education*, 20, (s. 393-419).
- GORDON, E.M. & BONILLA-BOWMAN, C. (1996). Can Performance-Based Assessments Contribute to the Achievement of Educational Equity? I B. Boykoff & D.P. Wolf (red.), *Performance-based Student Assessment: Challenges and possibilities*. Ninety-fifth Yearbook of the National Society for the study of Education. Chicago: The University of Chicago Press.
- GRUBER, H. LAW, L-C. MANDL, H. & RENKL, A. (1999). *Situated Learning and Transfer: Implications for Teaching*. I P. Murphy (red.), *Learners, Learning and Assessment*. London, Thousand Oaks och New Dehli: Paul Chapman Publishing in association with The Open University.
- GUSTAFSSON, B., Stigebrandt, E. & Ljungvall, R. (1982): *Den Dolda Läroplanen*. Stockholm: Liber Läromedel.
- HAERTEL, E. H. (1999). Performance Assessment and Education Reform. I *Phi, Delta Kappan*, Vol. 80 No. 9, (s. 67-3 – 679).
- HAMBLETON, R.K. (1990). Criterion-referenced Assessment in Education. I H. Walberg & G. Haertel (red.), *The International Encyclopedia of Educational Evaluation*. Oxford:Pergamon Press.
- HAMBLETON, R.K., SIRECI, S.G. (1997) Future Directions for Norm-referenced and Criterion-referenced achievement testing. *International Journal of Educational Research*; vol. 27 n5 (s. 379-93)
- HARGREAVES, A. (1989). *Curriculum and assessment reform*. London: Open University Press.
- HARGREAVES, A. (1998). *Läraren i det postmoderna sambället*. Lund: Studentlitteratur.

HARRIS, D. & BELL, C. (1994). *Evaluating and Assessing for Learning*. London: Kogan Page.

HARTMAN, N. J. (1998). *Vetenskapligt tänkande*. Lund: Studentlitteratur.

HERRNSTEIN, R. J. & MURRAY, C. (1994). *The Bell Curve: intelligence and class structure in American Life*. New York, London: Free Press.

HUSÉN, T. (1999). *Insikter och åsikter om utbildningsambället*. Stockholm: Gothia.

ISLING, Å. (1988). *Kampen för och emot en demokratisk skola*. Stockholm: Sober förlag.

JOHANSSON, J-E. (1999): *Några texter om pedagoger och pedagogik*. Manus

JONES, G.W. M. FL. (1999): The Impact of High-Stakes Testing on Teachers and Students in North Carolina in *Phi, Delta, Kappan* Vol. 81 No. 3, (s. 199-204).

JÖNSSON, I., TRONDMAN, M., ARMAN, G. & PALME, M. (1994). *Skola Fritid Framtid*. Lund: Studentlitteratur.

KOGAN, M. (1990). Fitting evaluation within the governance of education. I M. Granheim, M. Kogan & U.P Lundgren (red.), *Evaluation as Policymaking*. London: Jessica Kingsley Publishers.

KORP, H. (1996). *Samma prov blir olika – en studie om provets betydelse för elevers lärande*. C- och D-uppsats i pedagogik, Göteborgs Universitet.

KVALE, S. (1990). Evaluation and Decentralization of Knowledge. I M. Granheim, M. Kogan, & U.P. Lundgren (red.), *Evaluation as Policymaking*. London: Jessica Kingsley Publishers.

LAUVÅS, P. (1998). Hva skjuler seg bak metodespørsmålene? I B. Ljung & A. Pettersson (red.), *Perspektiv på bedømming av kunnskap*. Lärarhögskolan i Stockholm: Institutionen för pedagogik.

LAVE, J. & WENGER, E. (1991). *Situated learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

LOHMAN, D.F. (1997). Lessons from the History of Intelligence Testing I L.M. Anderson (red.), *Educational Testing and Assessment: Lessons from the Past, Directions for the Future..*

LUNDGREN, U.P. (1979). *Att organisera omvärlden*. Stockholm: Liber.

LUNDGREN, U.P. (1990). Evaluation and Decentralization of Knowledge. I M. Granheim, M. Kogan & U.P. Lundgren (red.), *Evaluation as Policymaking*. London: Jessica Kingsley Publishers.

LÜBCKE, P. (1988). *Filosoflexikonet*, s. 3 10. Stockholm: Bokförlaget Forum.

MABRY, L. (1999). Writing to the Rubric – Lingering effects of Traditional Standardized Testing on Direct Writing Assessment. I *Phi, Delta, Kappan*, Vol. 80 No. 9, (s. 673 -679).

MADAUS, G. F. & O'DWYER, L. (1999). A Short History of Performance Assessment – Lessons Learned. I *Phi, Delta, Kappan*, maj 1999 vol. 80 No 9 (s. 688-696).

MARTON, F. & BOOTH, S. (1997). *Learning and Awareness*. Mahwah och New Jersey: Lawrence Erlbaum Associates.

MARTON, F., HOUNSELL, D. ENTWISTLE, N. (1986: red.). *Hur vi lär*. Stockholm: RabËn Prisma.

MESSICK, S. (1993). Validity. I R. Linn (red.), *Educational Measurement*. Phoenix: American Council of Education and The Oryx Press.

MOORE, A. (1993). Gender and ethnocentricity in bilingual English classrooms. I P. Woods, & M. Hammersley (red.), *Gender and Ethnicity in Schools*. London: Open University Press/New York: Routledge.

MOSS, P.A. (1992). Shifting Conceptions of Validity. I *Educational Measurement: Implications for Performance Assessment. Review of Educational Research*, 62(3), (s. 229-258).

MÅHL, P. (1998). *Vad krävs nu? En bok om hur skolan kan se ut och fungera*. Stockholm: HLS Förlag.

NATIONALENCYKLOPEDIN (1995). Höganäs: Bra Böcker.

NEWMAN (1997). Authentic Assessment in Social Studies. I G. Phye (red.), *Handbook of Classroom Assessment – Learning, Adjustment and Achievement*. San Diego, London, Boston, New York, Sydney, Tokyo och Toronto: Academic Press.

NORSTEDTS SVENSKA UPPSALGSBOK OCH ORDBOK (1999). Göteborg: Norstedts.

NUTTALL, D. (1995). *Effective Assessment and the Improvement of Education*. London: The Falmer Press.

PERRENOUD, P. (1991) Towards a Pragmatic Approach to Formative Assessment. I P. Weston (red.) *Assessment of Pupil Achievement: Motivation and School Success*. Rapport från The Educational Research Workshop, som hölls i Liège 1989. Amsterdam/Lisse: Swets & Zeitlinger B.V.

PHYE, G. (1997). Classroom Assessment – A Multidimensional Perspective. I *Handbook of Classroom Assessment – Learning, Adjustment and Achievement*. San Diego, London, Boston, New York, Sydney, Tokyo and Toronto: Academic Press.

POLLARD, A. (1995) *The Social Construction of Pupil Identity*. London: Cassell.

POPKEWITZ, T. (1990). Some problems and problematics in the production of evaluations. I M. Granheim, M. Kogan & U.P. Lundgren (red.) *Evaluation as Policymaking*. London: Jessica Kingsley Publishers.

PUNTAMBENKAR, S. & DU BOULAY, A. (1999). Design of MIST – A System to Help Students Develop Metacognition. I P. Murphy (red.), *Learners, learning & assessment*. London, Thousand Oaks & New Dehli: Paul Chapman Publishing and The Open University.

QUELLMALZ, E. & HOSKYN, J. (1997). Classroom Assessment of Reasoning Strategies. I G. Phye (red.) *Handbook of Classroom Assessment – Learning, Adjustment and Achievement*. San Diego, London, Boston, New York, Sydney, Tokyo, Toronto: Academic Press.

RAMIEZ, A.L. (1999). Assessment-Driven Reform: The Emperor Still Has No Clothes. I *Phi, Delta, Kappan*, Vol. 81 No. 13 Nov 1999 (s. 204-08)

RAPPE, E. (1973): Från katekes till social fostran. Om folkundervisningen i Sverige från 1842års folkundervisningskommittÉ till 1906 års folkundervisningskommittÉ. I *Häftan för kritiska studier* 1973.

RAVEN, J (1991). Motivation, Effective Schooling and Assessment. I P. Weston (red.) *Assessment of Pupil Achievement: Motivation and School Success*. Rapport från The Educational research workshop, som hölls i Liège 1989. Amsterdam/Lisse: Swets & Zeitlinger B.V.

- RICHARDSON, G. (1999): *Svensk utbildningshistoria*. Sjätte upplagan. Lund: Studentlitteratur.
- ROMHEDE, R. (1999). Marknadsplats, myndighet eller mötesplats? I Alexandersson, M.(red.), *Styrning på villovägar*. Lund: Studentlitteratur.
- ROSÉN, M (1998). *Gender differences in patterns of knowledge*. Göteborg: Acta Gothoburgensis.
- ROWNTREE, D. (1977). *Att bedöma elever: hur kan vi lära känna dem?* Lund: Studentlitteratur.
- SAX, G. (1989). *Principles of Educational and Psychological Measurement and Evaluation* (3rd edition). Belmont, California: Wadsworth Publishing Company.
- SCB (2002). *Universitet och högskolor. Grundutbildning – Nybörjare, registrerade och examina 2000/01*. Sveriges officiella statistik, Statistiska meddelanden:. Serien Utbildning och forskning.
- SCB (2001). *Befolkningens utbildning*. Sveriges officiella statistik, Statistiska meddelanden:. Serien Utbildning och forskning.
- SCHÜLLERQUIST, U. (1996). Förskjutningen av svensk skolpolitisk debatt under det senaste decenniet. I Englund (red.) *Utbildningspolitiskt systemskifte?* Stockholm: HLS.
- SCRIVEN, M. (1991). Beyond Formative and Summative. I M.W. Mc Laughlin & D.C. Phillips (red.), *Evaluation and Education: At a quarter Century*. Chicago: Chicago University Press.
- SCRIVEN, M. (1994). Evaluation: Formative, Summative, Goal-free. I T. Husén, & N.T. Postlewaithe (red.), *The International Encyclopedia of Education*, 2nd edition, vol. 1. Oxford: Pergamon.
- Shepard, L. (2000). *The Role of Classroom Assessment in Teaching and Learning*. CSE Technincal Report.

SKOLVERKET (2001). *Bedömning och betyg – kommentarer med frågor och svar*. Stockholm: Liber distribution.

SKOLVERKET (1994). *Betygsboken: om betygsriterier i ett nytt betygssystem för gymnasieskolan och gymnasial vuxenutbildning*. Stockholm: Liber distribution.

SMITH, N.L. (1994). Evaluation Models and Approaches. I T. HusÈn & N.T. Postlewaithe (red.), *The International Encyclopedia of Education*, 2nd ed., vol. 1: 2101-2109, Oxford: Pergamon.

SOU 1992:94. *Skola för bildning*. Huvudbetänkande av läroplanskommittÈn. Stockholm: Allmänna förlaget.

STABERG, E-M. (1992). *Olika världar, skilda värderingar: hur flickor och pojkar möter högstadiets fysik och teknik*. Umeå: Akademiska avhandlingar vid Pedagogiska Institutionen.

STAKE, R. (1999). The Goods on American Education. I *Phi, Delta, Kappan*, Vol. 80 No. 9 (s. 668-72).

STIGGINS, R.J. (1998). Assessment, Student Confidence, and School Success. I *Phi, Delta, Kappan*, Vol. 81. No. 1, (s. 191-198).

SVINGBY, G (1998). *Utvärdering av grundskolan 1995: UG 95. Vad är rätt och rättvist?: Ungdomars val i etiska situationer som uttryck för demokratisk kompetens: samhällsorienterande ämnen, årskurs 9*. Stockholm: Statens Skolverk: Liber distribution.

SÄLJÖ, R. (1975). *Qualitative differences in learning as a function of the learners conception of the task*. Göteborg: Acta Universitatis Gothoburgensis.

SÄLJÖ, R. (1994). Minding Action. Conceiving of the world versus participating in cultural practices. *Nordisk pedagogik*, 14, (s. 71-80).

SÄLJÖ, R. (2000). *Lärande i praktiken – ett sociokulturellt perspektiv*. Stockholm: Bokförlaget Prisma.

TAYLOR, C. (1994). Assessment for Measurement or Standards: The Peril and Promise of Large Scale Assessment Reform. *American Educational Research Journal*, Summer 1994, Vol. 31 No. 2, (s. 231-262).

TESFAHUNEY, M. (1999) Monokulturell Utbildning i *Utbildning & Demokrati*, Vol. 8, Nr 3, (s. 65-84).

TORRANCE, H. (Red.) (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.

TORRANCE, H. (2001). *Educational Assessment in the Next Millennium: the challenge of postmodernism*. I G. Svingby & S. Svingby (red.), Konferensrapport från konferens om bedömning av kunskap och kompetens som hölls vid Malmö Högskola 17-19 nov. 1999.

UNENGE, G. (1992) *Romarna*. D-uppsats i pedagogik. Göteborgs Universitet.

UTBILDNINGSDEPARTEMENTET (1994). 1994 års Läroplan för de frivilliga skolformerna, Lpf 94. Stockholm: Fritzes.

UTBILDNINGSDEPARTEMENTET (1994). *Läroplan för det obligatoriska skolväsendet*, Lpo 94. Stockholm: Fritzes.

UTBILDNINGSDEPARTEMENTET (1998). *Läroplan för det obligatoriska skolväsendet, förskoleklassen och fritidsbemmet* (Lpo 94/98). Stockholm: Fritzes.

WEINER, G. & BERGE, B-M. (1994). *Kön och kunskap*. Lund: Studentlitteratur.

WILLIS, P. (1981). *Fostran till lönearbete*. Göteborg: Röda Bokförlaget.

WINTHER JÖRGENSEN, M. & PHILLIPS, L. (1999).
Diskursanalys som teori och metod. Lund: Studentlitteratur.

WOLF, D.P. & REARDON, S.Y. (1996). Access to Excellence Through New Forms of Student Assessment. I B. Boykoff & D.P. Wolf (red.), *Performance-based Student Assessment: Challenges and possibilities*. Ninety-fifth Yearbook of the National Society for the study of Education. Chicago: The University of Chicago Press.

WOLF, R.M. (1994). Evaluation, empirical framework for. I T. HuseËn & N.T. Postlewaithe (red.), *The International Encyclopedia of Education*, 2nd ed., vol. 1 (s. 2089-2097). Oxford: Pergamon.

WORTHEN, B.R. & VAN DUSEN, L.M. (1994). Evaluation, Nature of. I T. HuseËn & N.T. Postlewaithe (red.), *The International Encyclopedia of Education*, 2nd ed., vol. 1 (s. 2109-2119). Oxford: Pergamon.

ÅBERG, R. (Red.)(1992): *Social bakgrund, utbildning, livschanser*. Stockholm: Carlsson Bokförlag.

ÅLUND, A. (1996). Etniska ungdomar och den sociala krisen – Om utbildning, arbete och identitet. I E-S. Wallentin & C. Hultinger (red.), *Den mångkulturella skolan*. Lund: Studentlitteratur

Bilaga I

Svensk forskning om kunskapsbedömning mellan 1990 och 2000

För att få en bild av den aktuella forskningen i Sverige om kunskapsbedömning har jag sökt i LIBRIS, de svenska universitetsbibliotekens gemensamma katalog. I tabellen framgår vilka söktermer som har använts och hur många relevanta poster som respektive sökning har resulterat i. Kriteriet för en relevant post är att den ska representera en skrift som är publicerad i Sverige tidigast 1990 och har kunskapsbedömning som sitt huvudsakliga ämne.

SÖKTERMER	ANTAL RELEVANTA POSTER
Bedömning	33
Bedömning Utbildning	2
Bedömning Kunskap	5
Bedömning Undervisning	10
Bedömning Lärande	1
Bedömning Skola	1
Bedömning Prov	3
Bedömning Elev o Prov	136
Prov Utbildning	24
Prov Lärande	2
Prov Undervisning	24
Prov Elev	1

Avgränsning och kategorisering

De titlar, som klassificerats som relevanta, har sorterats under de sex rubrikerna "Forskning", "Allmän litteratur", "Utveckling av bedömningsmetoder", "Utvärderingar", "Prov" och "Oklassificerade". Till "Forskning" har förts rapporter från olika slags forskning om kunskapsbedömning, både sådan som är mycket lokal och tillämpningsorienterad, och sådan som är mera generell och teoretiskt inriktad. Till "Allmän litteratur"

räknas (universitets-) läroböcker om kunskapsbedömning och böcker, där frågor om kunskapsbedömning diskuteras på vetenskaplig grund. Beskrivningar av projekt som direkt syftar till att utveckla och förbättra specifika bedömningsinstrument har hänförs till kategorin "Utvecklingsarbete". Till "Utvärderingar" har förts sådana utvärderingar som till väsentlig del bygger på och utgörs av forskning om kunskapsbedömning. Faktiska prov och tester har sammanförts under rubriken "Prov". Titlar, som inte kan klassificeras i någon av de övriga kategorierna, eller som kan hänföras till flera kategorier och sådana där det är osäkert vad för slags skrift de betecknar är samlade under rubriken "Oklassificerade". I många fall har det av kataloginformationen inte varit möjligt att dra säkra slutsatser om vad för slags skrift det är fråga om och på vilket sätt den knyter an till ämnet kunskapsbedömning.

Kategorin "Prov", som är den i särklass största (exempelvis representerar 83 av 136 de relevanta poster, som sökningen på termen "prov" resulterade i, prov av olika slag), innehåller inga vetenskapliga skrifter och är därför inte intressant att undersöka. I kategorierna "Forskning" och "Allmän litteratur" representerar samtliga poster skrifter av intresse, och dessutom förekommer forskning om kunskapsbedömning under de båda rubrikerna "Utvärderingar" och "Utveckling av bedömningmetoder".

I det nedanstående presenteras samtliga titlar i de fem kategorier som kan innehålla forskning om kunskapsbedömning.

1. Forskning om kunskapsbedömning

1. Att studera och bli bedömd. Andersson, P. Linköping, 2000
2. Påverkar provprocessen och olika former av prov elevers lärande? Marcussen, E.-M. & Velander, Å. Luleå, 1999
3. Tankar om tal. Palmér, A. – Uppsala, 1999
4. Målrelaterade och normrelaterade prov. Wiberg, M. – Umeå, 1999
5. Principen och praktiken. Östlund-Stjärnegårdh, E. – Uppsala, 1999

6. Perspektiv på bedömning av kunskap. Konferensantologi Stockholm, 1998
7. Bedömning av kvalitet i matematikkunskaper. Nyström, P. – Umeå, 1998
8. Bedömning av kunskap i klinisk praktik. Jerreling, L. & Svenningsson, Agardh, I. – Växjö, 1997
9. I skärningspunkten mellan kvalitativ och kvantitativ vetenskap: en jämförelse mellan FN-soldaters testresultat från DORT och soldaternas militära betyg och omdömen
10. Högskoleprovet. Gustavsson, M. – Linköping, 1995.
11. Enkät till lärare i svenska vid gymnasieskolor med tvååriga linjer rörande det centrala provet – analys och kommentar. Elliot, I. & Eneskär, B. – Malmö, 1993
12. Grundläggande kunskaper i aritmetik. Englund, T. m fl – Stockholm, 1993
13. Kvalitativa aspekter på provkonstruktion. Helmstad, G. & Ekeblad, E. – Mölndal, 1992
14. Grundläggande kunskaper i aritmetik. Pettersson, A. & Santesson, AC. – Stockholm, 1992
15. Matematikprov och lärarprestation. Runesson, U. & Ekholm, M. – Linköping, 1992
16. Gymnasieelever om skolbetyg och högskoleprov. Askling, B. & Danielsson, C. – Linköping, 1993
17. Testträning. Henriksson, W. – Umeå, 1990
18. Gruppdifferenser och bias. Stage, C – Umeå, 1990
19. Könsskillnader på högskoleprovet våren 1990. Stage, C. – Umeå, 1990
20. Textmängdens betydelse för prestationen på läs-provet. Umeå, 1990

2. Allmän litteratur om kunskapsbedömning

1. Pedagogik för 2000-talet, Eigidius, H. – Stockholm, 1999
2. Svenska på prov. Nyström, C & Ohlsson, M. (Red.) – Uppsala 1999.
3. Portfolio. Ellmin, R. – Göteborg, 1999.
4. Vad krävs nu? Mähl, P. – Stockholm, 1998.

5. Portfoliometoden, Taube, K. – Stockholm, 1997
6. Utvärdering och bedömning i praktisk utbildning. – Linköping, cop. 1997
7. "Läraryrket". Allard, B. M fl – Stockholm, 1992
8. Betygsättning och utvärdering i skola och utbildning. Eigidius, H. – Malmö, 1992
9. Tånagelareans betydelse för kroppsmedelvärder. Ramstedt, K. – Umeå, 1992
10. Ett nytt betygssystem. Slutbetänkande från betygsberedningen (SOU) – Stockholm, 1992
11. Betyg och prov. Forneng, S. & Jansson, H. – Stockholm, 1991
12. Examination som kvalitetskontroll i högskolan. Lindberg-Sand, Å. & Askling, B. – Stockholm, 1991
13. Betyg men på vad? Måhl, P. – Stockholm, 1991

3. Utvärderingar av vetenskaplig karaktär och där forskning om kunskapsbedömning ingår som en väsentlig del.

1. Utvärdering av grundskolan 1995. Proven, kunskapen och undervisningen – Stockholm, 1998
2. Utvärdering av grundskolan 1995. Muntlig framställning – Stockholm, 1997
3. Den nationella utvärderingen av grundskolan våren 1992 – Stockholm, 1993
4. Den nationella utvärderingen av grundskolan våren 1992. Engelska – Stockholm, 1993
5. Göteborgsdiagnosen. Runesson, U. & Ågren, M. – Göteborg, 1991
6. Den nationella utvärderingen – Stockholm, 1991
7. Utvärdering av engelska i årskurs 5. Balke, G. m fl – Göteborg, 1990

4. Utveckling av bedömningsmetoder

1. Rätt eller fel? Andersson, K. Umeå, 1999
2. Från utprovning till färdigt prov. Hellekant, J. – Mälndal, 1997
3. Forskning kring tolv års centrala prov i svenska för treåriga linjer 1984 - 1995. Eneskär, B. Malmö, 1996
4. Utkast till en metod för bedömning av nationella prov i matematik. Ramstedt, K. – Umeå, 1994
5. Allt är relativt – Göteborg, 1991
6. Forskning kring tio års centrala prov i svenska för tvååriga linjer 1984- 1993. Eneskär, B – Malmö, 1994
7. Forskning kring treåriga linjers centrala prov i svenska under åren 1984- 1989. Eneskär, B. – Malmö, 1994
8. Nytt teoriprov 1990. Mattsson, H. – Umeå, 1990
9. Vad tycker provdeltagarna om högskoleprovet 1990-05-05? Wester-Wedman, A. – Umeå, 1990

5. Oklassificerade titlar

1. Bedömning av social kompetens vid rekrytering av mellanschefer. Blomdahl, M. — Luleå, 2000
2. Kunskapsutveckling hos elever med rörelsehinder. Malmqvist, J – Göteborg, 2000-08-10
3. Pedagogik för 2000-talet. Eigidius, H. – Stockhomm, 1999
4. Texten och inspirationskällan. Persson, A. – Uppsala, 1999.
5. Framgång måste ständigt vinnas på nytt. Högskoleverket – Stockholm, 1998
6. Provbanksförsök fysik A occ B samt matematik D Våren 1998. Boesen, J. M m fl – Umeå, 1998.
7. Acroski – en bedömningssport.. Bedömningarnas tillförlitlighet och relevans. Johansson, A. – Umeå, 1998
8. Rättvis rättning i nationella prov. Lindström, J-O. – Umeå, 1998
9. Värdekriterier i samband med bedömning av resultat i sångredovisnings-sammanhang. Norberg, B. – Örebro, 1998
10. Informationssökning med IT i ett pedagogiskt sammanhang. Svedberg, S. Hässelby, 1997.

11. Referensbinding i elevuppsatser. Öberg, H-S. – Uppsala, 1997.
12. Skriva debattartiklar i skolan- går det? Östlund-Stjärnegårh, E. – Uppsala, 1997
13. Att pröva skrivförmågan med nationella prov. Garne, B. – Uppsala, 1996.
14. Samspel och kommunikation. Garne, B. & Palmér, A. – Uppsala, 1996
15. Elektriska flickor och mekaniska pojkar. Ramstedt, K. – Umeå, 1996
16. Prov i modersmålet. Andersson, M. – Uppsala, 1995
17. Skrivandet, kursplanerna och läromedlen. Nyström, C. – Uppsala, 1996
18. Invandrare och test. Book, C. – Solna, 1995
19. På ny kurs i engelska – Göteborg, 1995
20. Gruppkillnader på centrala prov i fysik. Ramstedt, K. – Umeå, 1995
21. Svenska i utveckling – Uppsala, 1995
22. Vad sa' du fröken? Bedömning av tal och språk hos nybörjare. Alstam-Malcus, A. & Fritzell, M. – Göteborg, 1993
23. Bedömning av språklig medvetenhet hos förskolebarn och skolbarn. Magnusson, E. & Nauclér, K. – Löddeköpinge, cop. 1993
24. Bedömning av lärares pedagogiska skicklighet. Karlsson, C. – Linköping, 1991
25. Allt är relativt – Göteborg, 1991 Nytt prov. Önskemålsnämnden – Stockholm, 1991
26. Bedömning av kunskaper och färdigheter i bild, årskurs 2. Eklund, S. – Umeå, 1990
27. Barn lär barn skriva. Holmberg, L. & Marklund, K. Kalmar, 1990

Metodiska problem

Avsikten med att söka litteratur med begreppen ”bedömning” och ”prov” och kombinationer av dessa med begreppen ”utbildning”, ”undervisning”, ”skola”, ”lärande” och ”elev” som söktermer var att få fram de svenska skrifter som uttryckligen behandlar frågor om kunskapsbedömning. En följd av att använda denna princip är dock att endast sådana skrifter hittas, som har någon av dessa termer är registrerade som ämnesord eller i sin titel.

Inom exempelvis ämnesdidaktisk forskning eller forskning om undervisning och lärande inom yrkesutbildning och högre utbildning kan det antas finnas skrifter, som inte har dessa ord i sin titel eller som ämnesord, men ändå innehåller forskning om kunskapsbedömning. Å andra sidan representerar större delen av det totala antalet visade poster skrifter helt utan anknytning till kunskapsbedömning, inom områden som medicin, teknik och kvalitetsgranskning. Bland de poster som kan relateras till pedagogikämnet, representeras flera sådana skrifter där andra ämnen än kunskapsbedömning är i fokus och endast resultaten av kunskapsbedömning tas upp.

Inom exempelvis ämnesdidaktisk forskning är det ofta nödvändigt att studera vad eleverna kunde eller hur de begrep något innan undervisningen har ägt rum och efter. Ofta ingår också beskrivningar och diskussioner av de bedömningsmetoder som har använts. I många fall är man också intresserad av att undersöka elevernas tänkande och vilka strategier som eleverna använder när de lär sig. Då bedöms även kunskapsformer som processkunskap, strategisk kunskap, kognitiva strategier och metakognition. I specialpedagogisk forskning kan man vara intresserad av att bedöma kunskapsutvecklingen hos elever med olika slags funktionshinder eller andra svårigheter som påverkar deras förutsättningar att lära sig. I dessa fall är alltså frågor om kunskapsbedömning att betrakta som forskningsmetodiska frågor, snarare än forskningsfrågor. Forskningsfrågorna gäller istället sådant som elevernas lärande inom ett visst ämne och vilka problem som kan finnas, eller hur undervisningen kan utvecklas för att stödja elevernas lärande på ett optimalt sätt.

Också de nationella utvärderingarna av skolan bygger i stor utsträckning på resultat från kunskapsbedömningar. I dessa är det skolans kvalitet och kunskapsstandarden hos svenska elever som är i fokus. För att kunna undersöka detta måste emellertid både prov och bedömningskriterier utvecklas och detta utvecklingsarbete, som görs av forskare, är utgör en kvalificerad form av bedömningsforskning.

Intelligensforskning och andra former av forskning om människors mentala förmågor bygger också på resultat av bedömningar av människors prestationer, där de studerade förmågorna antas komma till uttryck. Sådan forskning hanterar naturligtvis också frågor om kunskapsbedömning i en mening¹.

En tillförlitlig kategorisering av litteratur kräver naturligtvis kännedom om innehållet i de ingående skrifterna. Nu baserar den sig i de fall då jag inte redan varit bekant med boken ifråga, på den knappa information som finns att tillgå i LIBRIS. En konsekvens av detta är att många titlar har placerats i kategorin "Oklassificerade", en annan att det troligen förekommer felsorterade titlar. Förhoppningsvis blir det möjligt för mig senare att undersöka innehållet i de böcker, som jag har presenterat här.

Även om denna översikt alltså måste betraktas som ytterst preliminär och även om samtliga av de oklassificerade titlarna skulle representera vetenskapliga skrifter inom området, så visar den att det i Sverige har publicerats anmärkningsvärt lite forskning om kunskapsbedömning. I synnerhet tycks det vara ont om forskning av "grundforskningskaraktär". Vid en snabb genomgång synes den vetenskapliga publikationen om kunskapsbedömning till den allra största delen vara inriktad mot pragmatiska frågeställningar, medan mera teoretiska frågor om exempelvis bedömningens sociala och ideologiska funktioner och dess pedagogiska konsekvenser är mera sällsynta som utgångspunkt för forskning och diskussion. Ett undantag som jag känner till är Per Anderssons nyligen utkomna doktorsavhandling Att studera och bedömas.

1 Förmåga och kunskap är visserligen två skilda begrepp. Med förmåga avses ens persons förutsättningar för att klara av olika saker och när det gäller kognitiv förmåga, eller begåvning, så har denna traditionellt betraktats som något inneboende hos individen, medfött och konstant. Kunskap betecknar istället resultatet av inläring, d v s något som individen själv har förvärvat i samspel med sin yttre omgivning. Senare års intelligensforskning har emellertid frångått den statiska uppfattningen av intelligens och man menar numera att intelligensen kan påverkas utifrån och att individers intelligens kan ökas genom att de får hjälp att utveckla sina kognitiva strategier. Därmed löses gränsen mellan kunskaps- och begåvningsbegreppet delvis upp.

SERIEN FORSKNING I FOKUS VID MYNDIGHETEN FÖR SKOLUTVECKLING

Är en skriftserie som etablerats för att möjliggöra utgivning av material som producerats med stöd eller på uppdrag av Myndigheten för skolutveckling.

Det gemensamma för skrifterna är att Myndigheten för skolutveckling gjort bedömningen att materialet är av intresse för t.ex. pedagogiskt yrkesverksamma, beslutsfattare och forskare.

Författarna svarar själva för innehållet och de ställningstaganden som görs.

Publikationerna finns också på hemsidan www.skolforskning.se och www.skolforskning.nu

Från och med 2003-03-01 övergick serien genom en omorganisation från Skolverket till Myndigheten för skolutveckling.

TIDIGARE FORSKNINGS- OCH KUNSKAPS- ÖVERSIKTER SOM PUBLICERATS:

Glenn Hultman & Cristina Hörberg

Kunskapsutnyttjande

Ett informellt perspektiv på hur kunskap och forskning används i skolan

Ingegerd Muncio

Genomförande

Vem tolkar beslut och vem ser till att reformer blir mer än ord?

Britt Hallerdt

Studieresultat och social bakgrund

- en översikt över fem års forskning

Kjell Granström & Charlotta Einarsson

Forskning om liv och arbete i svenska klassrum

- en översikt

Ingrid Pramling Samuelsson & Ulla Mauritzson

Att lära som sexåring

En kunskapsöversikt

Birgitta Sahlin

Matematiksvårigheter och svårigheter när det gäller koncentration i grundskolan

En översikt av svensk forskning 1990-1995

Erik Wallin

Gymnasieskola i stöpsleven – då nu alltid

Perspektiv på en skolform

Mats Börjesson

Om skolbarns olikheter

Diskurser kring ”särskilda behov” i skolan – med historiska jämförelse punkter

Hans Ingvar Roth

Den mångkulturella parken

- om värdegemenskap i skola och samhälle

Ulla Forsberg

Jämställdhetspedagogik

- en sammanställning av aktionsforskningsprojekt

Jens Pedersen

Informationstekniken i skolan

En forskningsöversikt

Peder Haug

Pedagogiskt dilemma:

Specialundervisning

Moira von Wright

Genus och text

När kan man tala om jämställdhet i fysikläromedel?

Tullie Torstenson-Ed & Inge Johansson

Fritidshemmet i forskning och förändring

En kunskapsöversikt

Thomas Tydén och Annika Andræ Thelin (RED)

Tankar om lärande och IT

En forskningsöversikt

Ulla Riis (RED)

IT i skolan mellan vision och praktik

En forskningsöversikt

Pia Williams, Sonja Scheridan och Ingrid Pramling Samuelsson

Barns samlärande

En forskningsöversikt

Monica Söderberg Forslund

Kvinnor och skolledarskap

En kunskapsöversikt

Mats Ekholm, Ulf Blossing, Gösta Kåräng, Kerstin Lindvall, Hans-Åke Scherp

Forskning om rektor

En forskningsöversikt

Ingemar Emanuelsson, Bengt Persson, Jerry Rosenqvist

Forskning inom det specialpedagogiska området

En kunskapsöversikt

Ulla Lind

Positioner i svensk barnpedagogisk forskning

En kunskapsöversikt

Karin Rönnerman

Vi behöver varandra

En utvärdering

- Irene Rönnerberg, Lennart Rönnerberg
Minoritetselever och matematikutbildning
En litteraturoversikt
- Maj Asplund Carlsson, Ingrid Pramling Samuelsson och Gunni Kärrby
Strukturella faktorer och pedagogisk kvalitet i barnomsorg och skola
- Karin Wallby, Synnöve Carlsson och Peter Nyström
Elevgrupperingar
En kunskapsöversikt med fokus på matematikundervisning
- Gunilla Härnsten
Kunskapsmöten i skolvärlden
Exempel från tre forskningscirkelrar
- Elisabet Öhrn
Könsmönster i förändring
En kunskapsöversikt om unga i skolan
- Rolf Helldin
Specialpedagogisk forskning
En kritisk granskning i ett omvärldsperspektiv
- Anna Klerfelt
Var ligger forskningsfronten
67 avhandlingar i barnpedagogik under två decennier, 1980-1999
- Louise Limberg, Frances Hultgren, Bo Jarneving
Informationssökning och lärande
En forskningsöversikt
- Jan-Eric Gustafsson, Eva Myrberg
Ekonomiska resursers betydelse för pedagogiska resultat
En kunskapsöversikt
- Björn Eriksson, Odd Lindberg, Erik Flygare och Kristian Daneback
Skolan – en arena för mobbning
En forskningsöversikt
- Eva R Fähræus, Lars-Erik Jonsson
Distansundervisning
Mode eller möjlighet
- Ingrid Nilsson
Fristående skolor
Internationell forskning 1985-2000
- Ingegerd Tallberg Broman, Lena Rubinstein Reich och Jeanette Hägerström
Likvärdighet i en skola för alla
Historisk bakgrund och kritisk granskning
- Birgit Lendahls Rosendahl och Karin Rönnerman
Handledning av pedagogiskt yrkesverksamma
– en utmaning för skolan och högskolan

Bernt Gustavsson

Vad är kunskap?

En diskussion om praktisk och teoretisk kunskap

Eva Johansson

Möten för lärande

Pedagogisk verksamhet för de yngsta barnen i förskolan

Gunnel Colnerud, Robert Thornberg

Värdepedagogik i internationell belysning

Per Andersson, Nil-Åke Sjösten och Song-ee Ahn

Att värdera kunskap, erfarenhet och kompetens

Perspektiv på validering

Lars Holmstrand och Gunilla Härnsten

Förutsättningar för forskningscirkel i skolan

En kritisk granskning

Joakim Ekman och Sladjana Todosijević

Unga demokrater

En översikt av den aktuella forskningen om ungdomar, politik och skolans demokrativärden