

Innehåll

Förord	7
1. Introduktion	9
Prov	10
Kunskap, förmåga och kunnande	11
Bakgrunden till dagens betygssystem	12
Sammanfattning	18
2. Användningsområden och vilseledning	20
Urval eller information? Norm- och målrelaterade prov	20
Summativt och formativt syfte	23
Tillåtna sätt att vilseleda: Gissa och bluffa	25
Gissning	25
Bluffning	27
Sammanfattning	29
3. Kvalitet	30
Direkta och indirekta mätningar	31
Mätfel	33
Slumpmässiga mätfel	34
Systematiska mätfel	36
Kopplingen mellan mätfel och kvalitet	37
Reliabilitet	38
Internt samband	39
Interbedömarreliabilitet	41
Intrabedömarreliabilitet	43
Parallellreliabilitet	46
Att analysera reliabilitet	47

KOPIERINGSFÖRBUD!

Detta verk är skyddat av upphovsrättslagen! Kopiering är förbjuden, utöver lärares rätt att kopiera för undervisningsbruk enligt avtal med Bonus Presskopia och den mycket begränsade rätten till kopiering för privat bruk. Den som bryter mot lagen om upphovsrätt kan åtalas av allmän åklagare och dömas till böter eller fängelse upp till två år samt bli skyldig att erlagga ersättning till upphovsman/rättsinnehavare.

© 2021 Jonathan Wedman
Författare: Jonathan Wedman
Förläggare: Ida Larsson
Lektor: Elswig Patriksson
Illustrationer: Izabella Balcerzak
Omslagsbild: HybridGraphics/stock.adobe.com
Språkgranskning, sättning och formgivning omslag: Vulkanisterna AB

Första utgåvan, första tryckningen
Framställd av Vulkan.se
Tryckt i Europa, 2021
ISBN 978-91-89335-41-7

Validitet	49
Innehållsbevis, del 1: Vad läraren bör mäta	52
Innehållsbevis, del 2: Vad läraren väljer att mäta	55
Innehållsbevis, del 3: Vilken bedömningsmetod läraren använder	56
Relationsbevis	56
Strukturbevis	58
Processbevis	61
Konsekvensbevis	62
Integration av bevis till ett sammanhållet validitetsargument	63
Sammanfattning	64

4. Bedömningsmetoder 66

Skriftliga prov: Stängda och korta frågor	68
Skriftliga prov: Essäfrågor	71
Praktiska prov	74
Muntliga prov	77
Sammanfattning	79

5. Taxonomi för lärande 80

Kognitiva nivåer och processer	82
Minnas	83
<i>Känna igen</i>	83
<i>Komma ihåg</i>	84
Förstå	84
<i>Tolka</i>	84
<i>Exemplifiera</i>	84
<i>Klassificera</i>	85
<i>Sammanfatta</i>	85
<i>Dra slutsatser</i>	85
<i>Jämföra</i>	86
<i>Förklara</i>	86
Tillämpa	86
<i>Genomföra</i>	87
<i>Applicera</i>	87

Analysera	88
<i>Särskilja</i>	88
<i>Organisera</i>	88
<i>Tillskriva</i>	89
Värdera	89
<i>Kontrollera</i>	89
<i>Kritisera</i>	90
Skapa	90
<i>Generera</i>	91
<i>Planera</i>	91
<i>Producera</i>	91
Sammanfattning	92

6. Bedömningsplanering 93

Skolverkets tre frågor	93
Bedömningsplanen	95
Steg 1: Identifiera självständiga delar av kunskapskraven ...	96
Steg 2: Identifiera verben i de självständiga delarna	99
Steg 3: Koppla kunskapskraven till det centrala innehållet .	100
Steg 4: Koppla kunskapskraven till bedömningsmetoder ...	102
Steg 5: Koppla kunskapskraven till kognitiva processer ...	102
Steg 6: Organisera bedömningsplaneringen i en tabell	104
Sammanfattning	106

7. Provkonstruktion 108

Provspecifikationer	108
Provspecifikationer för stängda och korta frågor	110
Provspecifikationer för essäfrågor, praktiska prov och muntliga prov	115
Frågekonstruktion	118
Riktlinjer för stängda och korta frågor	120
<i>Innehåll</i>	121
<i>Bedömningsform</i>	124
<i>Stil</i>	124
<i>Stam</i>	125
<i>Alternativen</i>	127

Riktlinjer för essäfrågor	131
<i>Konstruktion</i>	131
<i>Utvärdering innan eleverna tar provet</i>	134
<i>Utvärdering efter att eleverna har tagit provet</i>	135
Riktlinjer för praktiska prov	136
Sammanfattning	138
8. Rättning och bedömning	140
Rättning eller bedömning?	140
Bedömningsanvisningar	143
Analytiska och holistiska bedömningsanvisningar	145
Generella och provspecifika bedömningsanvisningar	148
Att skriva bedömningsanvisningar: Steg för steg	149
Bedömningsanvisningar som en process	152
Nyttan med bedömningsanvisningar: Tre perspektiv	153
Sammanfattning	155
9. Bedömningsskalor	156
Allmänt om bedömningsskalor	156
Raka poäng	159
Nivåpoäng	162
Betygsbeteckningar	164
Procent	168
Sammanfattning	169
Ordlista	171
Referenser	185

rekommendationer ur ett mättekniskt perspektiv. Det vill säga, att gå emot dessa rekommendationer leder för det mesta till försämrade mätningar men inte till rena lagbrott. Detta kan vara bra att veta för den reflekterande läsaren som kanske råkat avvika från någon rekommendation här och där.

Jag skriver också, på olika ställen i boken, att Skolverket råder eller avråder från vissa saker, och att de använder vissa termer, med mera. Jag för dock inte Skolverkets talan, varken i någon formell eller informell bemärkelse, så dylika formuleringar ska betraktas som mina tolkningar av Skolverkets texter, och ska inte ses som officiella uttalanden från Skolverket. Jag står självklart bakom mina tolkningar och jag anser mig kunna motivera dem på goda grunder, men läsaren är fri att bilda sin egen uppfattning av dem om denne så vill.

Prov

Det allra första som ska adresseras i den här boken är ordet *prov*. Det kanske inte är så konstigt med tanke på att *Prov* också är bokens titel. Att ordet prov dessutom förekommer i olika former och böjningar cirka 800 gånger i den här boken gör det också högst relevant att tydliggöra mer exakt vad det är jag menar.

Det finns lite olika definitioner av prov, bland annat beroende på sammanhanget, så till den här boken har jag valt att formulera en definition som överensstämmer i så hög grad som möjligt med det som Skolverket kallar för *bedömningssituation*. Eftersom den här boken är riktad främst mot bedömning med ett summativt syfte är definitionen också vinklad mot en summativ användning av provresultat. Definitionen av prov såsom ordet används i denna bok, är följande:

Ett prov är en situation där syftet dels är att eleven uppvisar någon typ av prestation som en lärare kan observera, och där syftet dels är att läraren ska fälla ett värderade omdöme om denna elevprestation, utifrån kriterier som tas upp i kursens eller ämnets kunskapskrav, vilka har tolkats av läraren och konkretiserats i ett facit eller en bedömningsanvisning.

Meningen är måhända lång och otymplig men den täcker in det som behöver täckas in: Att ett prov kan förekomma i olika former, eller situationer, att det handlar om olika typer av elevprestationer, att lärare måste kunna observera prestationerna, att lärare ska värdera prestationen, att värderingen ska vara relevant utifrån kunskapskraven, att kunskapskraven först måste ha tolkats av läraren och att det som bedömningen i praktiken sker mot är facit eller bedömningsanvisningar, beroende på vilken bedömningsmetod och -form som användes i den aktuella situationen.

Dess otymplighet till trots har jag funnit denna definition användbar vid diskussioner och reflektioner om prov. Oavsett om du som läsare håller med om detta hoppas jag ändå att den hjälper till att göra boken och dess innehåll lättare att greppa.

Kunskap, förmåga och kunnande

Förutom ordet prov är det en del andra ord som återkommer genom boken. Tre av de mest relevanta ur ett bedömningsperspektiv är *kunskap*, *förmåga* och *kunnande*. När förmåga nämns handlar det om praktiska förmågor och färdigheter. För att något ska vara en förmåga räcker det inte med teoretisk kännedom om hur något ska genomföras utan det kräver också att eleven kan utföra något i praktiken.

Interbedömarreliabilitet kan undersökas på mer eller mindre avancerade sätt, och ett smidigt sätt att undersöka den i skolan är att beräkna överensstämmelsen mellan bedömare. Rent konkret inleds det med att två lärare bedömer elevprestationerna i en hel klass. Efter det går man igenom, elev för elev, hur respektive lärare bedömde prestationen och noterar samtidigt om lärarna är överens eller inte. Avslutningsvis räknar man om det i procent så om två lärare exempelvis är överens om en viss bedömning för 21 av 30 elever så motsvarar det att de är överens i 70 % av fallen, vilket ger en interbedömarreliabilitet om 0,70.

Gränsen för vad som räknas till ”tillräckligt” när det kommer till interbedömarreliabilitet brukar vara just 0,70 i utbildnings-sammanhang. I exemplet ovan skulle tillförlitligheten därför betraktas som acceptabel. Om man tänker sig exemplet ovan, med 21 av 30 elever som blivit bedömda likadant, kan sambedömning med fördel användas för att på nytt bedöma de nio eleverna vars prestationer värderats olika av de båda lärarna. På så sätt kan även dessa elever få ett resultat som lärarna är överens om.

Intrabedömarreliabilitet är av naturen svårt att undersöka för den egna klassen eftersom det kräver en konsensus om vilka elevsvar eller -prestationer som är att betrakta som likvärdiga, och först efter det kan reliabiliteten undersökas. Problemet är ju då att en lärare som har bedömt sina elever redan har värderat vilka elevsvar som är att betrakta som lika och inte vilket gör att analysen blir metodologiskt haltande.

Parallellreliabilitet undersöks med en vanlig Pearson-korrelation för linjära samband. Den analysmetoden kommer vi inte gå igenom här men det är en välkänd metod som finns beskriven i åtskillig litteratur. Om korrelationen är tillräckligt hög kan de båda provversionerna anses vara jämförbara, eller utbytbara, och det är försvarbart att använda dem som

ordinarie prov och omprov. I fall parallellreliabiliteten är låg tyder det på att provversionerna inte mäter samma kunskap och då minskar likvärdigheten i bedömningsunderlaget som läraren försöker samla in. En låg parallellreliabilitet innebär också att åtminstone en av de båda provversionerna mäter något annat än den valda delen av kunskapskraven.

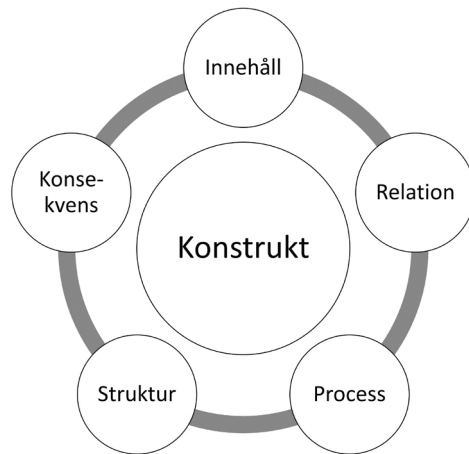
Validitet

Validitet är den vetenskapliga benämningen för det som Skolverket tidigare kallade *trovärdighet* och numera kallar *giltighet*. Det finns flera vetenskapligt grundande ramverk för validitet och de följer olika steg får att nå samma mål. Vissa ramverk använder termen *validering* (processfokus) medan andra använder *validitet* (produktfokus). I grund och botten är dock syftet med samtliga ramverk samma sak: Att ta reda på om ett prov mäter rätt sak på rätt sätt.

I den här boken presenteras den moderna synen på validitet, vilken är resultatet av ett gemensamt utvecklingsarbete mellan American Educational Research Association (AERA), American Psychological Association (APA) och National Council on Measurement in Education (NCME) (2014). Den moderna synen på validitet innefattar fem typer av validitetsbevis som binds samman till ett validitetsargument om provets kvalitet, och som har sin utgångspunkt i provets konstrukt och den tänkta användningen av provresultatet (Figur 12).

Dessa fem typer av validitetsbevis är innehålls-, relations-, struktur-, process-, och konsekvensbevis. För att undersöka validitet krävs alltså att läraren, antingen ensam eller som del i ett större lärarlag, bedömer provet professionellt utifrån de fem typerna av validitetsbevis och integrerar dessa tillsammans med

konstruktet och den planerande användningen för att formulera ett validitetsargument till stöd för provet.



Figur 12. Modell av den moderna validitetsteorin.

Det som är viktigt att komma ihåg här är att det är först när läraren har formulerat ett sådant här validitetsargument, och alltså har utvärderat provets giltighet, som det går att avgöra om provet bör utgöra en del av ett allsidigt och ändamålsenligt bedömningsunderlag vid betygssättning. Att ingå i ett lärarlag är att föredra när validitet ska undersökas, eftersom olika lärarperspektiv är värdefulla vid bedömningsplanering och passar väl in på vad Skolverket kallar *samsyn*.

Validitet påminner om, men ska inte blandas ihop med, Skolverkets användning av begreppet *ändamålsenlighet*. Den stora skillnaden mellan de två är att validitet alltid är kopplat till tolkningen och användningen av en specifik mätning medan ändamålsenlighet är ett bredare begrepp som syftar på hur användbart något är för ett visst syfte.

När man skriver eller talar om ”provets validitet” (vilket jag även gör i den här boken) så är det en förenkling av att skriva ”den utsträckning i vilken provresultaten har tolkats och använts på ett korrekt och rättvisande sätt i förhållande till provets syfte och avsedda användningsområde”. Ändamålsenlighet, å andra sidan, kan även vara relevant att använda när man diskuterar exempelvis bedömningsmetoder, kognitiva nivåer, utrustning eller andra aspekter av bedömningssituationen. Resten av kapitlet ägnas åt de fem validitetsbevisen:

1. Innehållsbevis (Bevis baserat på innehåll)
 1. Vad läraren bör mäta
 2. Vad läraren väljer att mäta
 3. Vilken bedömningsmetod läraren använder
2. Relationsbevis (Bevis baserat på relationer till andra mätningar)
3. Strukturbevis (Bevis baserat på intern struktur)
4. Processbevis (Bevis baserat på svarsprocesser)
5. Konsekvensbevis (Bevis baserat på konsekvenser)

Den första typen av validitetsbevis, *innehållsbevis*, är också den mest omfattande och är därför indelad i tre underkategorier. Innehållsbevis innefattar dels att koppla provet till det tänkta innehållet, men också att koppla det vidare till undervisningen och konstruktet samt att utvärdera valet av bedömningsmetod och kvaliteten på frågorna i provet. Den här bevisstypen handlar alltså både om det som läraren väljer att mäta, det som läraren bör mäta och hur läraren väljer att mäta det, samt hur allt detta ska kopplas ihop genom så kallad *konstruktiv länkning* för att stödja validitetsargumentet för provet.

8. Rättning och bedömning

Under tiden som lärare konstruerar sina prov bör de även utforma instruktioner till hur provresultaten ska värderas. Dessa instruktioner brukar kallas för *facit* när det handlar om stängda frågor och vissa kortvarsfrågor, och kallas för *bedömningsanvisningar* för övriga bedömningsmetoder. Beroende på vad som används så kallas det att lärare antingen rättar eller bedömer elevprestationerna. De här kapitlet fokuserar på dessa båda begrepp, och kanske framförallt på det större och mer komplexa av de två, nämligen bedömning.

Rättning eller bedömning?

Två olika begrepp används för att beskriva vad som ibland verkar vara samma sak, nämligen *rättning* och *bedömning*. Men det finns en viktig skillnad mellan de två. Rättning är i de allra flesta fall något objektivt medan bedömning är subjektiv. Annorlunda uttryckt kan man säga att en person som helt saknar ämnesrelevanta kunskaper ändå kan rätta ett prov, om provet efterfrågar svar som antingen är rätt eller fel och om det finns en tydlig rättningsmall att följa.

Om stängda provfrågor har använts kan rättning dessutom ske automatiskt och maskinellt av en dator. Flera personer som alla saknar ämnesrelevanta kunskaper kan dessutom rätta samma elevsvar och komma fram till samma slutsats om elevens prestation. Deras rättning av provet skulle alltså bli likvärdig

även om de inte kände till ämnet som provet var avsett att mäta.

I motsats till detta finns bedömning. Bedömning är ett subjektivt och professionellt utlåtande som kräver att bedömaren både är insatt i ämnet och dessutom vet exakt vad provet är ute efter att mäta. Bedömningar är till sin natur mer subjektiva och mindre likvärdiga än rättningar och kräver därför tydliga och utförliga bedömningsanvisningar för att minska subjektiviteten i mätningen. På grund av deras subjektiva natur, och den lägre överensstämmelsen mellan lärare som bedömer än lärare som rättar, riskerar de att medföra en lägre tillförlitlighet och därmed en sämre validitet i underlaget som samlas in. Mätningar som kräver bedömning ska därför användas med försiktighet och enbart när någon del av kunskapskraven kräver det.

I praktiken bör rättning och bedömning betraktas som varsin ände på en kontinuerlig skala, istället för som två separata kategorier. När ett prov ska poängsättas är det sällan helt objektivt och sällan helt subjektivt utan i regel lite av varje (Figur 18). Stängda frågor är som bekant så nära den objektiva sidan av skalan man kan komma men det finns små inslag av subjektivitet även där, särskilt om eleverna har besvarat provet på papper istället för på dator. Då kan vissa elever ha markerat otydligt, suddat eller kryssat över något av sina svar, alternativt markerat två eller flera svarsalternativ även om instruktionerna endast efterfrågade ett.



Figur 18. De olika bedömningsmetodernas placering på en skala som går från objektiv rättning till subjektiv bedömning.