

# Vetenskapsrådet borde ändra sitt betygssystem

Curie den 24 januari 2024

Att mäta och jämföra kvalitet i forskning är välkänt svårt. När fler söker medel än det finns pengar till behövs dock rutiner för att betygssätta och rangordna ansökningarna.

Vetenskapsrådet betygsätter kvaliteten enligt fyra baskriterier: "Projektets vetenskapliga kvalitet", "Nytänkande och originalitet", "Sökandes kompetens", "Genomförbarhet". Utifrån dessa betyg ska ett femte "Sammanfattande betyg" sättas. Betygsskalan är sjugradig: 1=Dålig, 2=Svag, 3=Bra, 4=Mycket bra. 5=Mycket bra till utmärkt, 6=Utmärkt, 7=Enastående. Genomförbarheten bedöms dock enligt en tregradig betygsskala.

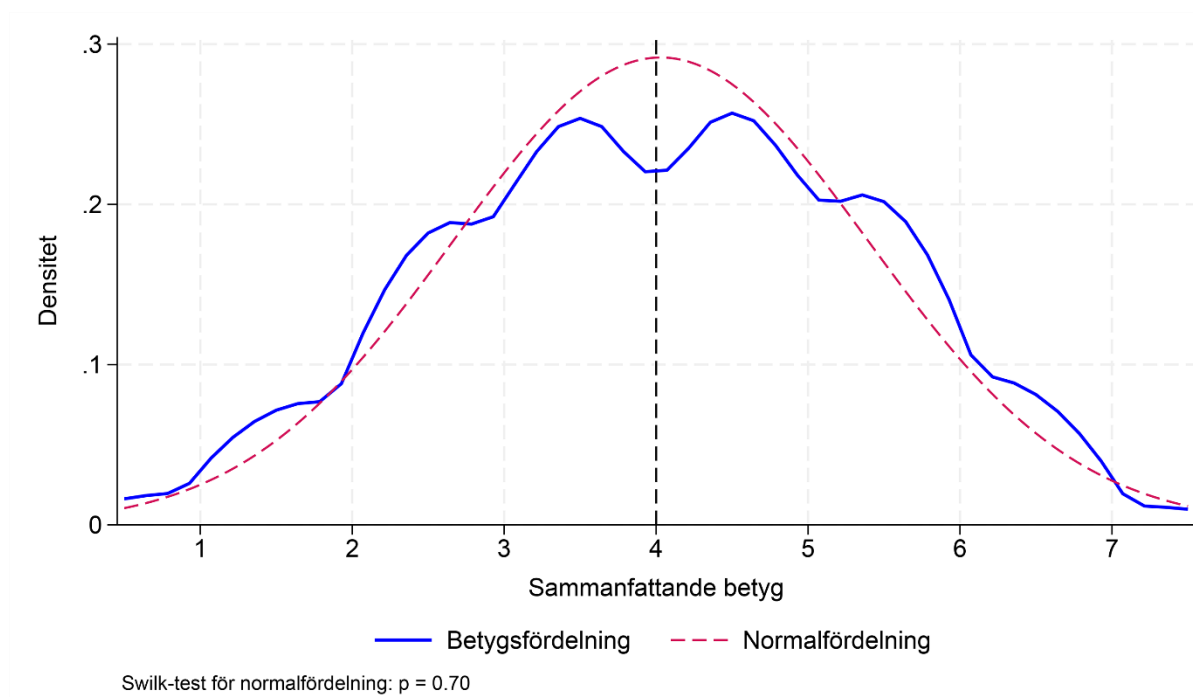
Tyvärr lider Vetenskapsrådets betygssystem av ett antal problem som påverkar finansieringsprocessen. Det har blivit uppenbart för mig när jag bedömt ansökningar i flera av Vetenskapsrådets utlysningar inom samhällskunskap och humaniora, både som ledamot och ordförande för olika beredningsgrupper.

## Problem1: Betygsskalan är asymmetrisk med ett alltför positivt mittenbetyg

Den sjugradiga betygsskalan får återkommande kritik. Det kanske största problemet är att skalan inte är symmetrisk kring en centrerad, neutral kvalitetsnivå. I stället är skalan avsiktligt skevt fördelad där mittenbetyget är "Mycket bra". Syftet är att öka möjligheten till urskiljning inom toppgruppen.

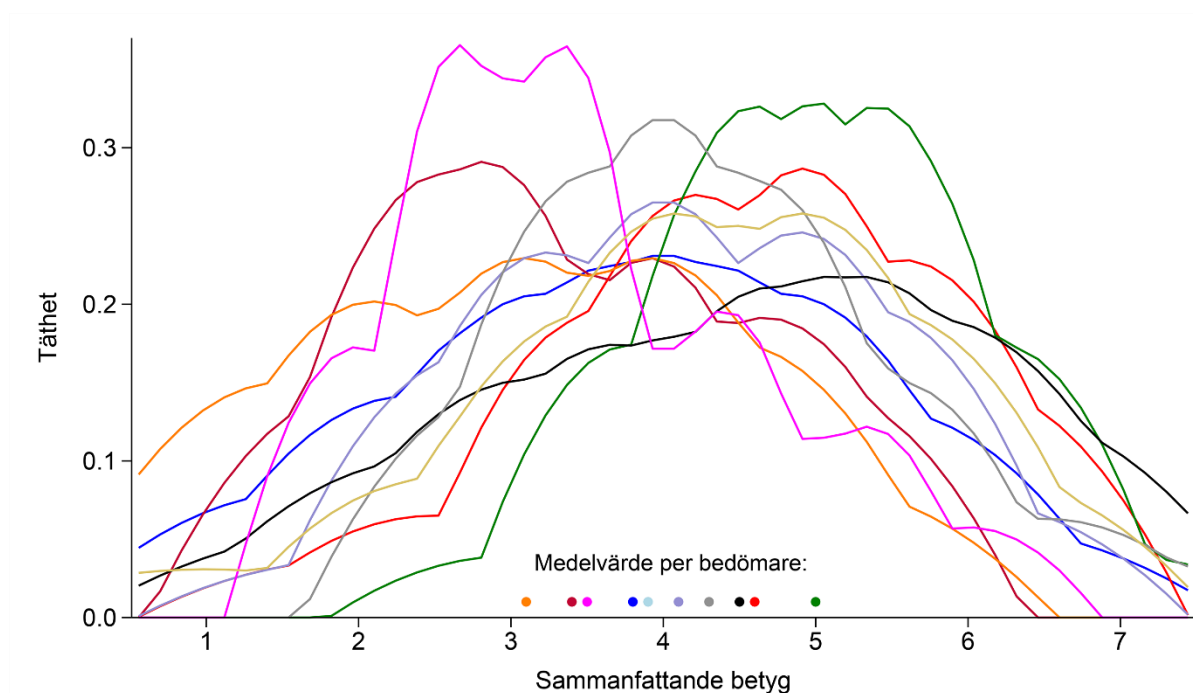
Men många bedömare tolkar ändå betygsskalan som vore den symmetrisk kring ett neutralt mittenbetyg, vilket skapar obalans och risk för godtycke. För att illustrera problemet använder jag betygsdata från utlysningen "Internationell postdok" inom samhällskunskap, utbildningsvetenskap och humaniora.

**Figur 1: Fördelning av bedömarnas totalbetyg på samtliga ansökningar**



Det finns givetvis skillnader mellan de olika bedömarnas betygssättning. Figur 2 visar de enskilda bedömarnas fördelning av de slutbetyg som deras ansökningar har fått. Figuren visar på en relativt stor spridning i betygsfördelning mellan bedömarna, trots att det finns ett överlapp eftersom varje ansökan bedöms av tre bedömare. Prickarna i botten visar bedömarnas respektive medelbetyg. I denna utlysning skiljer det hela två betygssteg mellan högsta och lägsta bedömarna.

**Figur 2: Fördelning av bedömarnas totalbetyg, per bedömare**



Att bedömare sätter olika betyg inget konstigt. Harmoniseringen vid bedömningsmötet är tänkt att hantera detta. Men om skillnaderna blir stora av skäl som inte beror på bedömarens läggning eller ansökningarnas egenskaper utan på betygsskalans konstruktion uppstår problem.

Diskussionerna vid de möten jag har varit på antyder att spännvidden i hög grad förklaras av tolkningsskillnader av betygsskalan. Och vid den tidpunkten är skadan redan skedd. Det är nämligen så att sållningen som tar bort cirka hälften av ansökningarna från vidare behandling görs efter första betygssättningen, och före harmoniseringsmötet. Ansökningar som har råkat få "lågbedömning" löper större risk att sållas ut än om de hade bedömts av "högbedömning".

Betygsskalorna som statliga forskningsfinansiärer i andra länder använder liknar Vetenskapsrådets, förutom på en punkt: deras mittenbetyg är neutralare.

Nationella forskningsfinansiärers betygsskalor i följande länder:							
Betyg* Finland	Frankrike	Nederländerna	Norge	Storbritannien	Sverige	USA	
9		Excellent					
8		Very good-Excellent					
7		Very good	Exceptional	Exceptional	Outstanding		
6	Outstanding	Excellent	Good-Very good	Excellent	Excellent	Excellent	
5	Excellent	Very good	Good	Very good	Very good	Very good-Excellent	Excellent
4	Good	Good	Moderate-Good	Good	Good	Very good	Very Good
3	Fair	Fair	Moderate	Fair	Fair	Good	Good
2	Poor	Weak	Subst.-Moderate	Weak	Weak	Weak	Fair
1	Insufficient	Poor	Substandard	Poor	Poor	Poor	Poor

## Problem 2: Betygsskalan har för få steg

Ett annat problem är att Vetenskapsrådets betygssteg är för få. Detta verkar även myndigheten anse eftersom de föreslår att bedömargrupperna lägger till extra kvalitetsnivåer utöver de som anges i anvisningarna. Specifikt handlar det om att bedömarna ges möjlighet att addera en kvalitetsdimension i tre nivåer, svag-medel-stark, till totalbetyget.

Anledningen till att Vetenskapsrådet inför fler betygssteg är helt enkelt för att erfarenheten visar att betygsskalan har för få steg. Alltför många ansökningar hamnar på samma betyg i samtliga bedömningsdimensioner.

## Förslag till lösning på problem 1 & 2: Symmetrisk betygsskala 1–10 eller 1–20

Mitt förslag är att Vetenskapsrådet inför en ny betygsskala skala som är symmetrisk 1–10 eller 1–20, där 1 är Mycket dåligt och 10 eller 20 är Extremt (ouppnåeligt) bra. Mittenbetygen motsvarar en neutralt godtagbar nivå.

Symmetriska skalor likt dessa används hela tiden och förstås av alla inom alla discipliner. Skalan 1–20 används i svensk grund- och gymnasieskola när bokstavsbedömning räknas om till sifferbetyg.

### Problem 3: Både delbetyg och sammanfattande betyg skapar inkonsekvens

Varje ansökan får fyra delbetyg och ett femte totalbetyg som ska bygga på de fyra första. Här uppstår flera problem. Ett är att det inte finns någon tydlig vägledning om hur sammanvägning ska ske. I praktiken ska bedömarna sätta vikter på de olika delkriterierna, men dessa vikter kan skilja sig åt betydligt mellan olika bedömare.

Vid det avslutande bedömningsmötet uppmanar Vetenskapsrådet dessutom gruppen att skapa explicita vikter, eller åtminstone rangordna mellan delkriterierna. Detta för att kunna rangordna ansökningar som fått samma slutbetyg även efter att de tre extrabetygen tillämpats.

Men inget av detta beskrivs tydligt i Vetenskapsrådets instruktioner till bedömarna. Det skapar osäkerhet och risk för godtycke. Risk för strategiskt beteende kan också uppstå. Min erfarenhet visar att erfarna bedömare har ett försteg eftersom de på förhand förstår den oproportionerligt stora betydelsen av totalbetyget, medan alla nya bedömare av naturliga skäl inte har uppnått denna förståelse.

### Förslag till lösning: Gör om totalbetyget till ett vägt genomsnitt av delkriteriernas betyg

En relativt enkel lösning vore att det sammanvägda totalbetyget görs om till att vara en funktion av delkriteriernas betyg. Här måste tydliga vikter bestämmas på förhand, vilket Vetenskapsrådet borde göra eftersom kriterierna och deras innehåll ju redan är formulerade. Ett exempel vore att sätta vikterna 0,35 på både vetenskaplig kvalitet ("Kvalitet") och nytänkande och originalitet ("Nytänkande"), 0,2 på den sökandens meriter ("Merit") och 0,1 på "Genomförbarhet", vilket ger följande beräkning av totalbetyget:

$$\text{Totalbetyg} = 0,35 * \text{Kvalitet} + 0,35 * \text{Nytänkande} + 0,2 * \text{Merit} + 0,1 * \text{Genomförbarhet}$$

Sammanfattningsvis har jag i denna text argumenterat för att en ny, symmetrisk betygsskala 1–10 eller 1–20 kring ett neutralt mittenvärde borde ersätta den betygsskala som Vetenskapsrådet använder idag. Totalbetyget bör sättas som ett vägt medelvärde av de olika delkriterierna och inte som idag vid sidan av delbetygen. Dessa förändringar skulle bidra till ökad tydlighet, konsekvens och stärkt rättssäkerhet i myndighetens finansieringsbeslut.

**Daniel Waldenström**