

EQUA

Bedömersamstämmighet och
resultat för svensklärarna

Vad är bedömersamstämmighet?



- Procentuell samstämmighet
 - Fullständig, angränsande
 - Värden som tar hänsyn till slumpen
- Samvariation
 - Korrelationer
- Stränghet
 - Medelvärden

Procentuell samstämmighet

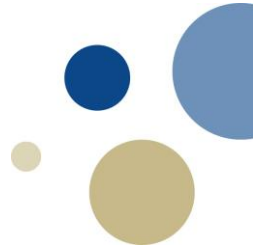


Förtest

	F	E	D	C	B	A
F	0	1	0	0	0	0
E	0	1	0	0	0	0
D	0	1	0	0	1	0
C	0	1	0	0	1	0
B	0	2	0	1	0	0
A	0	0	0	0	1	0

Två bedömare i engelskgruppen (skala: betyg): : lägst på förtest, högst på eftertest

Procentuell samstämmighet



Förtest

	F	E	D	C	B	A
F	0	1	0	0	0	0
E	0	1	0	0	0	0
D	0	1	0	0	1	0
C	0	1	0	0	1	0
B	0	2	0	1	0	0
A	0	0	0	0	1	0

Procentuell samstämmighet: 10 %
 Angränsande samstämmighet: 60 %
 Kappa-värde: -0,08
 Kappa-värde angränsande: 0,14

Eftertest

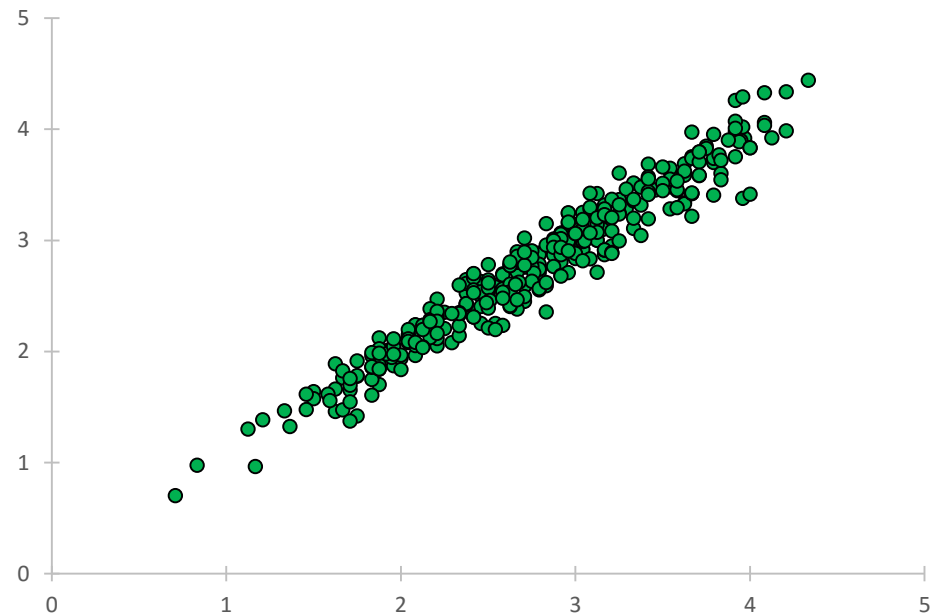
	F	E	D	C	B	A
F	0	0	0	0	0	0
E	0	2	0	0	0	0
D	0	0	1	0	0	0
C	0	0	0	1	0	0
B	0	0	0	1	0	0
A	0	0	0	0	1	4

Procentuell samstämmighet: 80 %
 Angränsande samstämmighet: 100 %
 Kappa-värde: 0,72
 Kappa-värde angränsande: 0,89

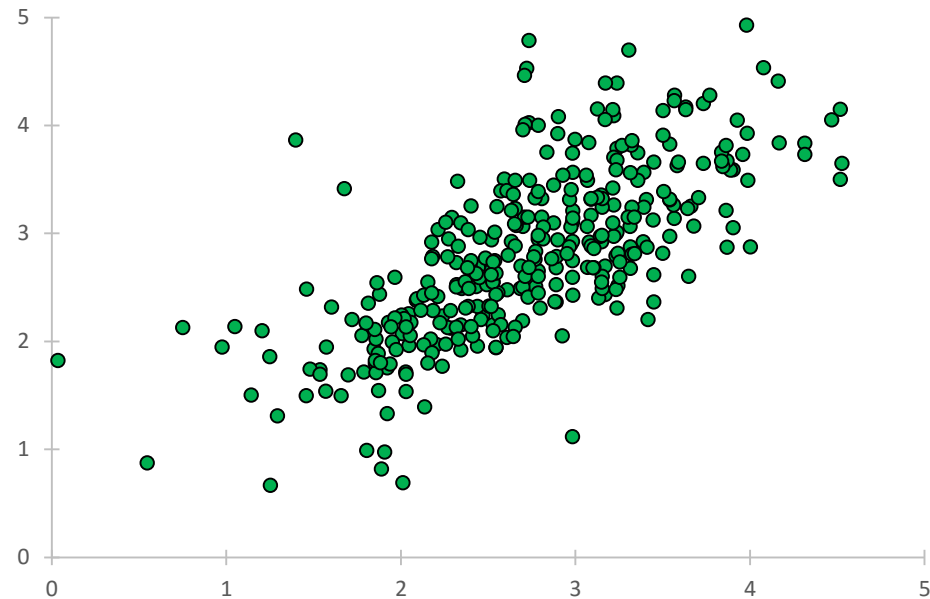
Två bedömare i engelskgruppen (skala: betyg): lägst på förtest, högst på eftertest

Korrelation/samvariation (t.ex. "ICC")

$r = ,97$



$r = ,71$



**,70: "rock-bottom minimum"
(McNamara, 2000)**

Stränghet

- Duva/hök



Svenskgänget - design



Förttest

- Individuell bedömning
- Bedömning av 30 elevtexter
- Fyra «aspekter» (1–4) + betyg (1–6)

Sambedömningsgrupp (4 pers.)

- Möte efter sommaren
- Diskussion/konsensus

Bedömarträning (8 pers.)

- Återkoppling: stränghet
- Föreläsning om bedömning
- Bedömningsträff med «facit»

Eftertest

- Individuell bedömning
- Bedömning av 30 elevtexter (6 ankar)
- Fyra «aspekter» (1–4) + betyg (1–6)

Frågor

- Blir grupperna mer samstämmiga?
- Är det någon skillnad mellan grupperna?



Svenskgänget – Kappa och procent samstämmighet

Betyg: kappa

	Förtest	Eftertest	
Sambedömning	0,24	0,27	↑
Bedömarträning	0,16	0,21	↑

Aspekter: procent

	Förtest	Eftertest	
Sambedömning	53,5 %	54,0 %	↑
Bedömarträning	47,8 %	52,5 %	↑



Samvariation: sambedomning



	Förtest		Eftertest		
	N	ICC	N	ICC	
Innehåll	29	,60	30	,59	↓
Struktur	29	,65	30	,50	↓
Språk	29	,75	30	,65	↓
Korrekthet	29	,69	30	,63	↓
Betyg	26	,67	30	,69	↑

Samvariation: bedömarträning



	Förtest		Eftertest		
	N	ICC	N	ICC	
Innehåll	20	,49	29	,61	↑
Struktur	20	,50	30	,56	↑
Språk	20	,57	30	,63	↑
Korrekthet	20	,56	30	,56	→
Betyg	20	,54	30	,67	↑

Svenskgänget - stränghet



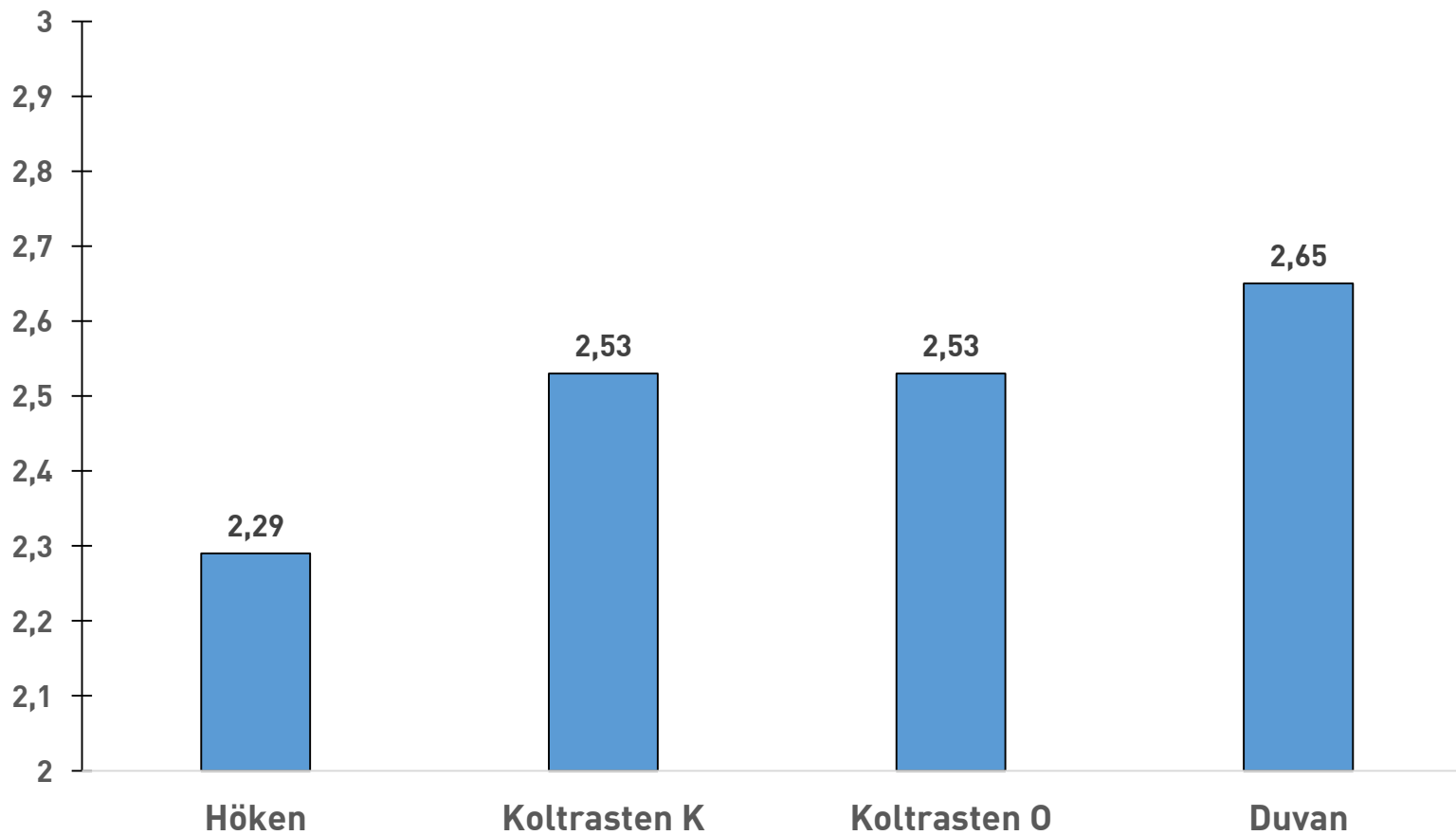
G-index

	Förtest	Eftertest
Sambedömning	2,2	2,7
Bedömarträning	1,6	3,6

Sambedömningsgänget: förtest



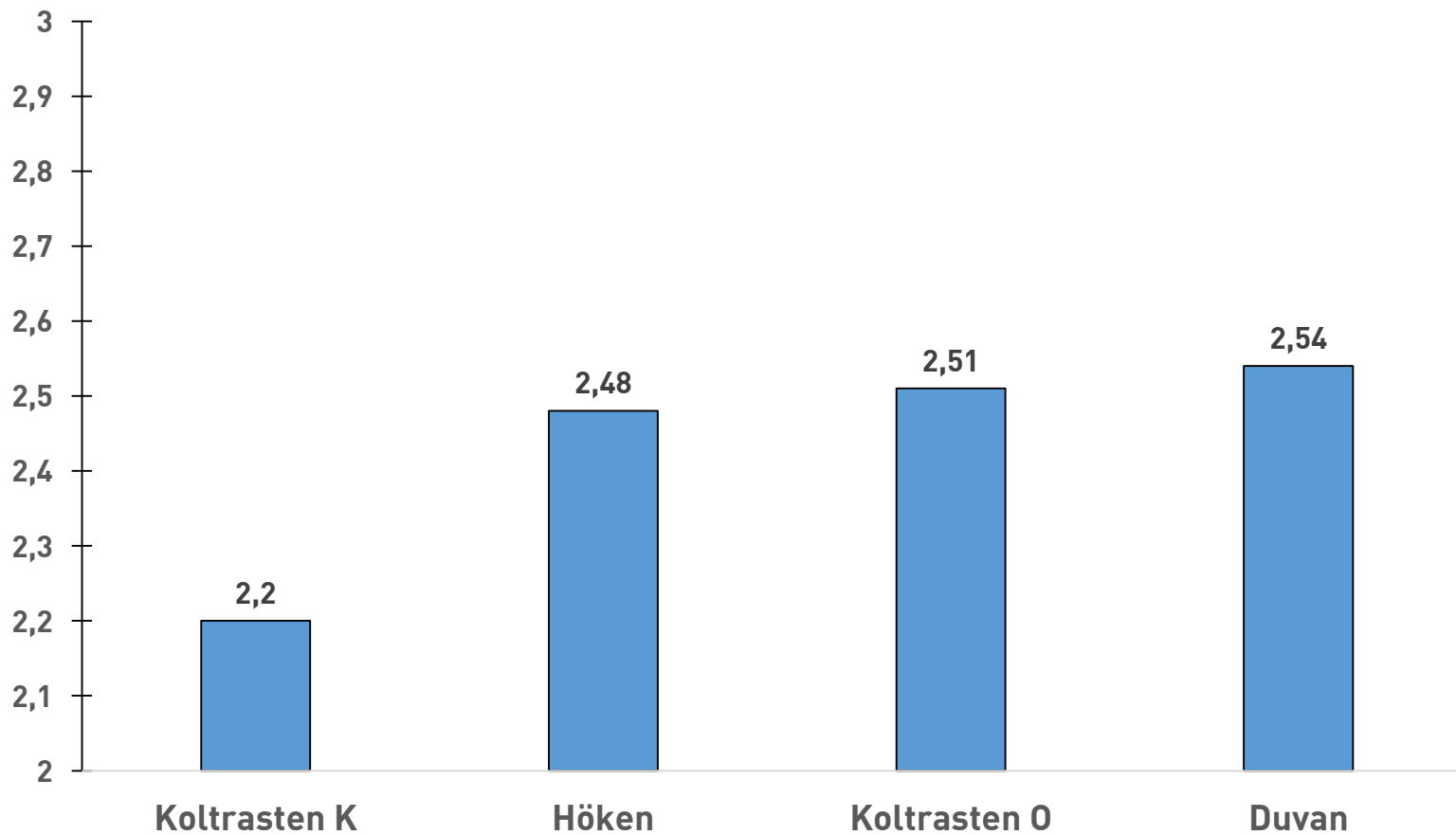
Sambedömning: pre-test



Sambedömningsgänget: eftertest



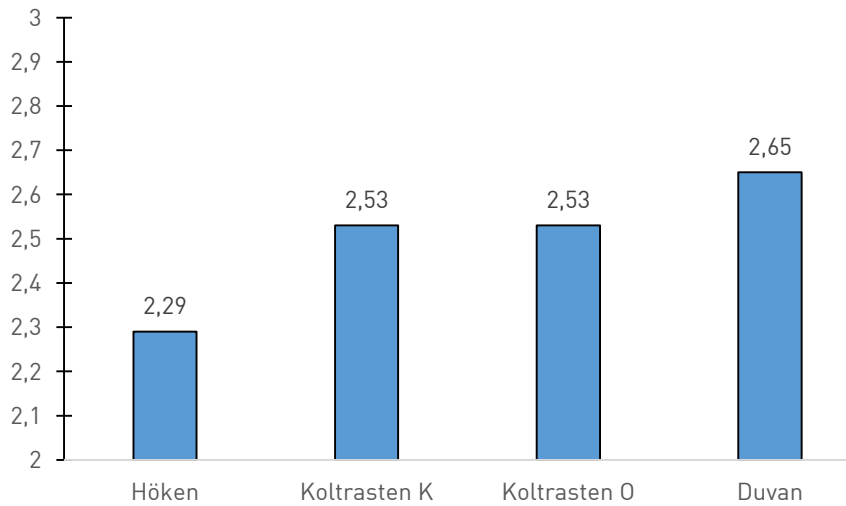
Sambedömning: post-test



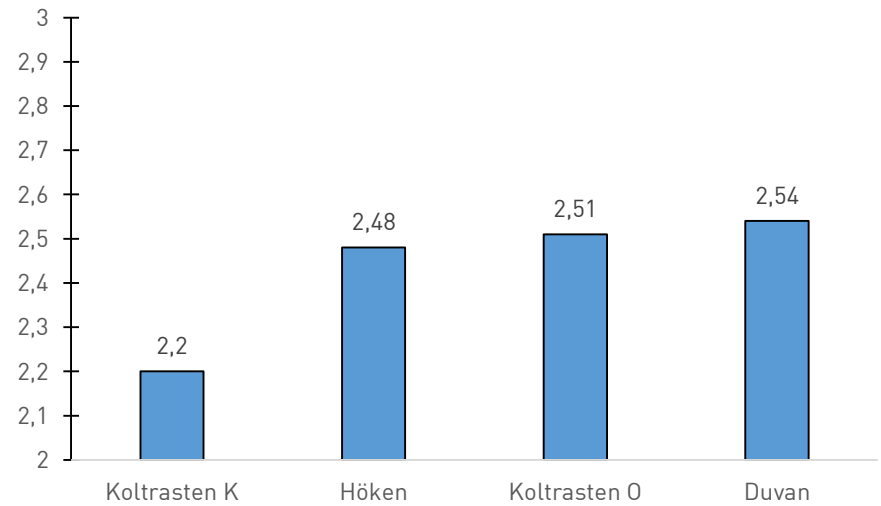
Förtest/eftertest



Sambedömning: pre-test



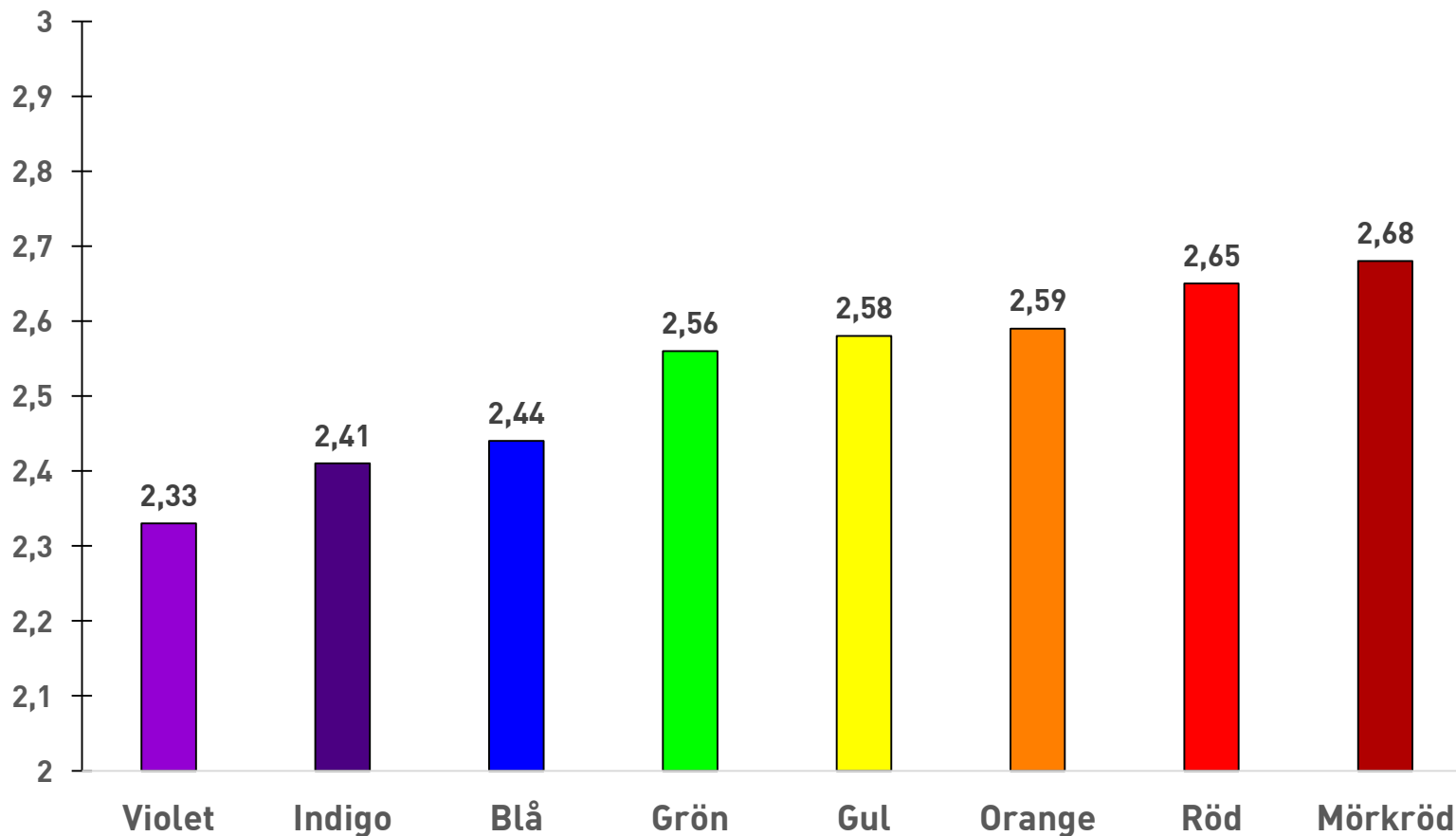
Sambedömning: post-test



Bedömarträning: förtest



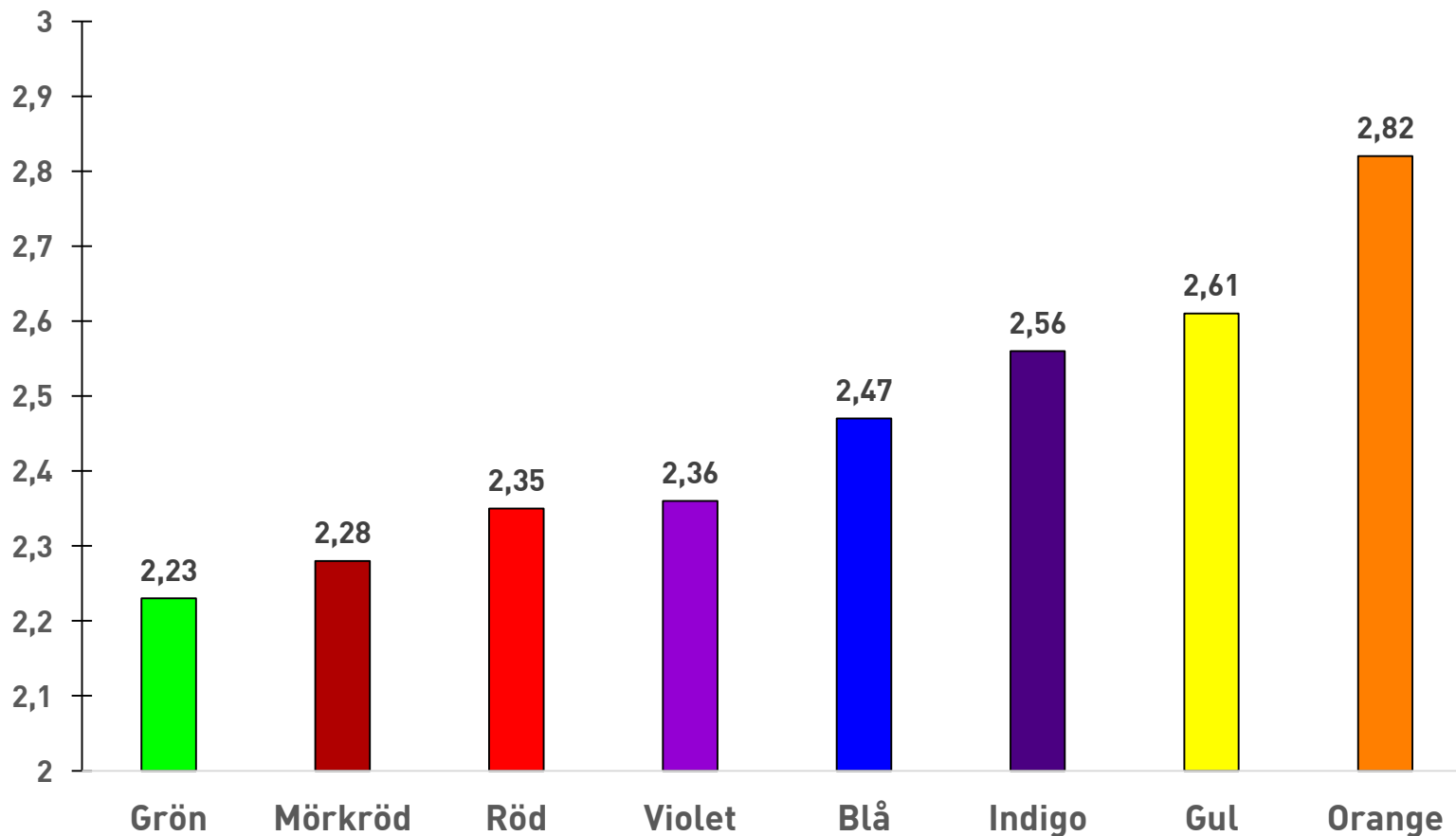
Bedömarträning: pre-test



Bedömarträning: eftertest



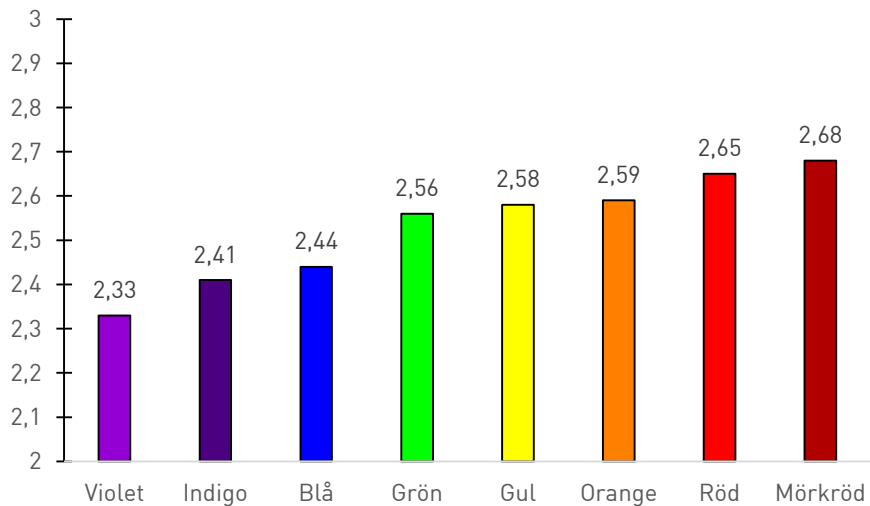
Bedömar-träning: post-test



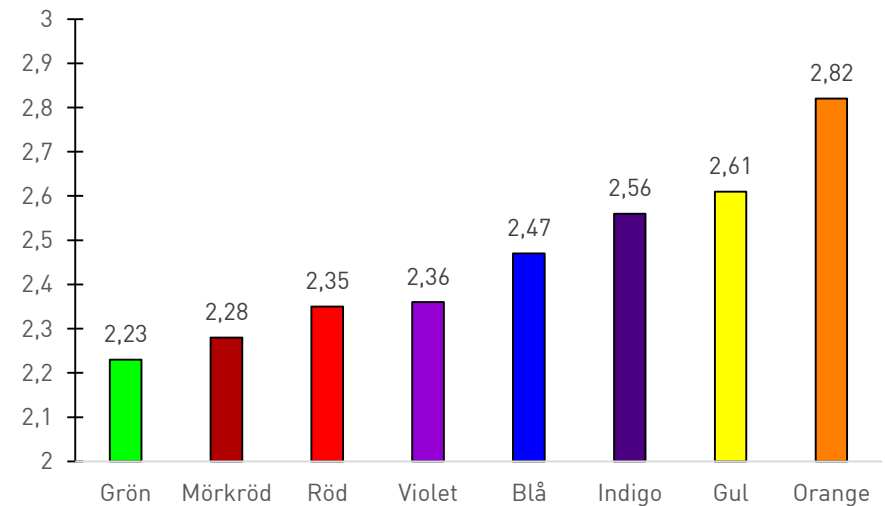
Förtest/eftertest



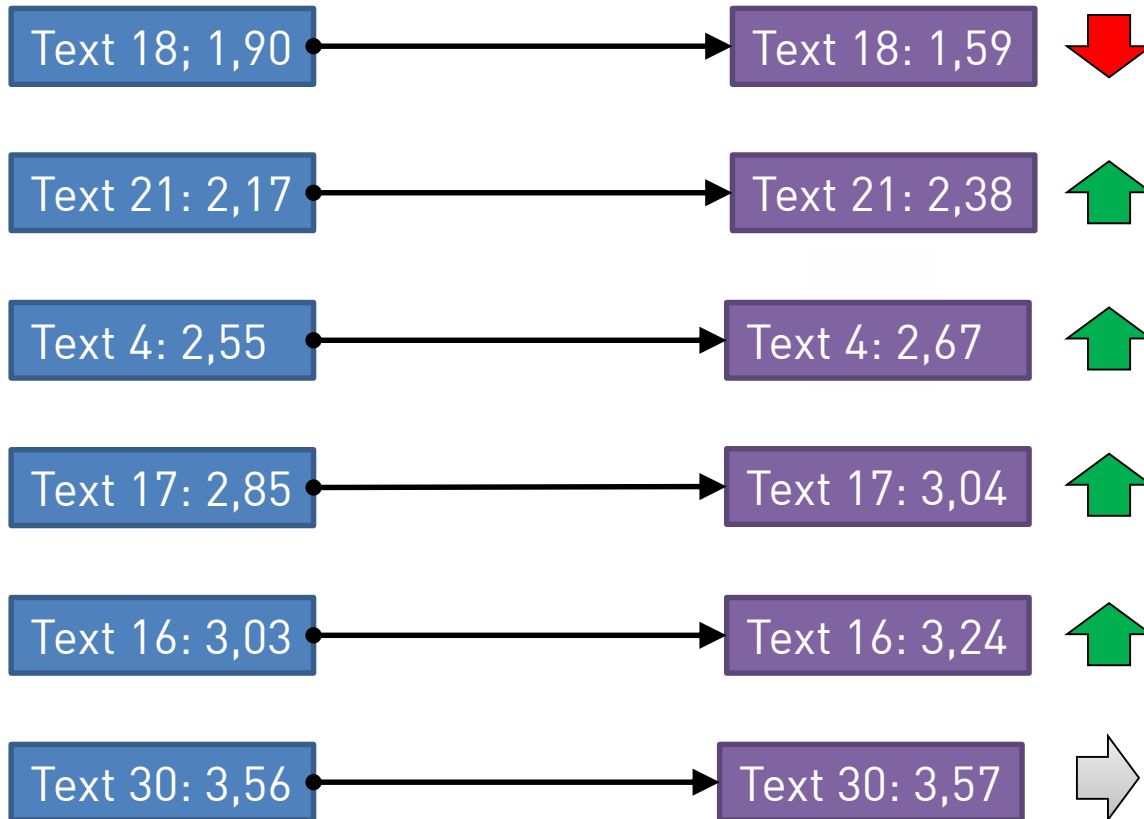
Bedömarträning: pre-test



Bedömar-träning: post-test



Hur gick det med ankartexterna?



Konklusioner för svenskänganget



- Som grupper: mycket stabila – jfr ankartexter
- Båda grupperna blir betydligt mer samstämmiga
 - Både **fokus på bedömning** och **träning fungerar**
- Återkopplingen till bedömarträningsgänget får konsekvenser (förändrad rangordning)
- Men, även om allt blir bättre:
 - **Ingen av grupperna når tillfredsställande reliabilitetsresultat**
 - **Elever löper större risk att stöta på bedömare som är oense, än han/hon har chans att stöta på bedömare som är «utbytbara»**
 - **Konsekvenser för likvärdigheten**



Kontakt

Gustaf B. Skar

gustaf.b.skar@ntnu.no

+47 916 93 654