



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

UHR

Att. Anders Ahlstrand

2026-09-01

Redovisning till UHR från projektgruppen för samordning och utveckling av prov i samhällskunskap¹ för medborgarskapsprov

Projektet MBP vid GU/IPS, grupperna för samhällskunskap och samordning

Föreliggande redovisning till UHR tar sin utgångspunkt i frågor som ställts till projektet för provutveckling i samhällskunskap respektive samordning vid Göteborgs universitet, Institutionen för pedagogik och specialpedagogik. Frågorna relaterar till den rapport som UHR ska leverera till Regeringskansliet (Utbildningsdepartementet) senast 30 september 2026. Rapporten bygger på det reviderade uppdrag vad gäller medborgarskapsprov som levererades i februari 2026. I detta beskrivs införandet av provet i tre steg (tidigare två):

- (1) Uppgifter i samhällskunskap (prototypuppgifter) att genomföras/prövas ut i aug. 2026 (med sikte på ett reguljärt delprov i samhällskunskap under 2027; se vidare s.8 ff, 'Samhällskunskapsprovet specifikt').
- (2) Uppdaterad tidplan och kostnadsberäkning för prov i hör- och läsförståelse i svenska för genomförande senast 1 okt 2027.
- (3) Analys av förutsättningarna för hur delproven i skriftlig respektive muntlig färdighet ska utformas, genomföras och bedömas; dessutom tidplan för införande av dessa delprov samt kostnadsberäkning.

¹Vi väljer i detta dokument att behålla beteckningen 'samhällskunskap', som är den som används officiellt, detta trots att det bredare ordet 'samhällsorientering' bättre speglar det som åsyftas i uppdraget och även i utbildningsmaterialet *Sverige i fokus*.



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Den här föreliggande rapporten har fokus på utprövningsprovet i samhällskunskap, som genomfördes den 15 aug 2026. Den inleds dock med ett avsnitt av mera principiell art med fokus dels på utprövning, dels på standardsättning, dvs. fastställande av kravgräns för ett godkänt resultat. Orsaken till detta är att de principer och procedurer som kännetecknar dessa förfaringssätt har otvetydig bäring på det fortsatta arbetet, såväl innehållsligt som tidsmässigt.

Steg 1 i införande av medborgarskapsprov

Steg 1 i införande av medborgarskapsprov innebär att delprovet som rör kunskaper om det svenska samhället provas ut i en adekvat grupp, dvs. sökande av svenskt medborgarskap, den 15 augusti 2026. Det är avgörande för nästa steg i införandet av ett medborgarskapsprov att kvaliteten i denna utprövning är så hög som möjligt. Här följer ett resonemang om varför antalet provtagare bör vara mycket stort, minst 500, och varför det är viktigt att gruppen är rimligt representativ.

Först vill vi kort beskriva ömsesidigheten mellan två typer av underlag: antalet uppgifter i ett prov och antalet deltagare som genomför provet. Många uppgifter ger ett bättre underlag för att tolka och fatta beslut om deltagarnas prestationer. På motsvarande sätt ger många deltagare ett bättre underlag för att bedöma hur väl varje uppgift fungerar. Med andra ord bidrar antalet uppgifter främst till säkrare bedömningar av deltagarnas prestationer, medan antalet deltagare bidrar till säkrare skattningar av uppgifternas kvalitet. Vår uppskattning är att antalet uppgifter bör vara minst 40 för att reliabiliteten ska vara tillräcklig² utan att arbetsbelastningen för provtagarna blir för stor, särskilt med tanke på att flera provdelar i svenska ska läggas till.

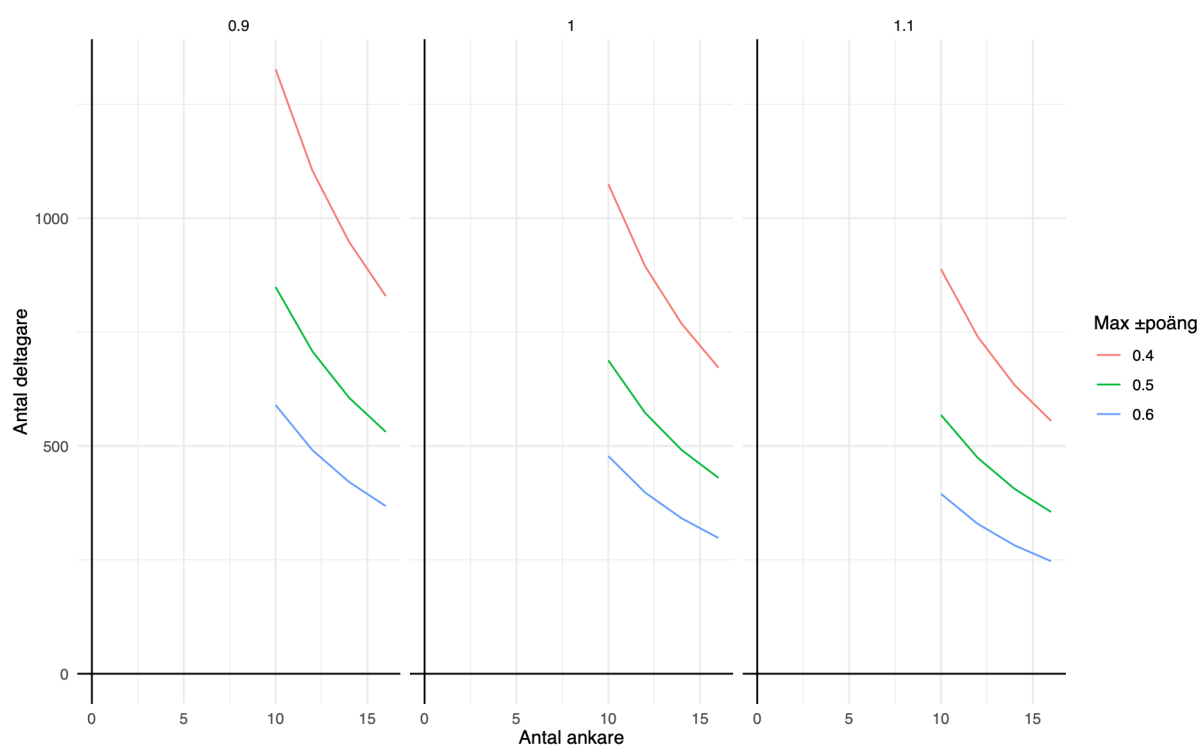
Noggrannheten i skattningarna av de enskilda uppgifternas svårighetsgrad och diskriminering (egenskap som skiljer prestationer åt) ökar således med antalet deltagare. Skattningarna behöver vara tillräckligt tillförlitliga för att kravgränssättning med hög kvalitet ska kunna genomföras. Därutöver är noggrannhet i ekvivalering - statistisk jämförelse och likställning -

² DeMars, 2010



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

mellan provversioner inte endast beroende av antalet deltagare, utan också av antalet uppgifter, kvaliteten hos dessa och antalet ankaruppgifter (identiska och gemensamma uppgifter i de båda provversionerna, resten olika). Med utgångspunkterna 40 uppgifter i provet, olika stora felmarginaler³ runt godkänthetsgränsen (G-gränsen), olika antal ankaruppgifter och några hypotetiska kvalitetsskattningar hos uppgifterna går det att simulera det ungefärliga antal deltagare som behövs.



Figur 1. Simulering av relationer mellan olika faktorer som påverkar kvaliteten vid ekvivalering mellan provversioner.

Figuren visar bland annat hur många provdeltagare som behövs för att uppnå en viss noggrannhet vid kravgränsen för godkänt resultat med utgångspunkt i tre olika scenarier för kvaliteten hos de i proven ingående uppgifterna. Längst till vänster i figuren har uppgifterna en genomsnittlig diskriminering (kvalitet) om 0,9, medan uppgifterna i mitten har 1 och

³ Standardfelet är särskilt viktigt vid G-gränsen eftersom små mätfel kan ändra klassificeringen av G och U. Ett resultat långt från gränsen påverkas sällan av ett litet fel. Ett resultat nära gränsen kan däremot hamna på olika sidor av gränsen beroende på ett mycket litet fel i poäng, ekvivalering eller skattning.



GÖTEBORGS UNIVERSITET INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

uppgifterna till höger har värdet 1,1 (realistiska värden att uppnå). Med krav på maximal felmarginal om $\pm 0,4$, $\pm 0,5$ respektive $\pm 0,6$ erhåller vi de röda, gröna och blå linjerna för 10 till 16 ankaruppgifter. En rimlig målsättning skulle kunna vara en genomsnittlig diskriminering hos uppgifterna på 1 och maximal felmarginal om $\pm 0,5$ för G-gränsen, vilket illustreras av den gröna linjen i mitten av figuren. Vid exempelvis 500 deltagare skulle det i så fall vara nödvändigt med ungefär 14 ankaruppgifter av de 40.

När det gäller representativitet skiljer sig gruppen som genomfört utvärderingsprovet sannolikt från hela populationen sökande av svenskt medborgarskap. Detta kan komplicera arbetet med att utvärdera de i provet ingående uppgifterna. Exempelvis kan det vara så att urvalet första gången, om t.ex. kravet på att alla andra villkor, inte minst dokumenterade svenskkunskaper, ska vara uppfyllda, presterar bättre än den genomsnittliga populationen som är aktuell för att genomföra medborgarskapsprovet. Det skulle kunna leda till så kallade takeffekter, d.v.s. att en stor andel av dem som genomfört provet får mycket höga poäng och att flera av uppgifterna framstår, utan att behöva vara det i hela populationen, som alltför enkla. Dessa takeffekter gör att uppgifterna kan uppvisa lägre mätbarhet i utvärderingsprovet 2026. I värsta fall kan detta leda till svårigheter att likställa svårighetsgraden mellan olika versioner av provet. Av denna anledning är det rimligt att eftersträva en större variation av svårighetsgrad mellan *items* som ingår i utvärderingsversionen av provet, än vad som kan bli fallet i framtida versioner av provet. Detta kan dock innebära ökad osäkerhet runt G-gränsen. Den statistiska analysen behöver därför kombineras med ett extra noggrant kvalitativt kravgränssättningsarbete (se vidare nedan).

Ovanstående resonemang innebär således att det, förutom uppgifter av god kvalitet, är viktigt med så många deltagare som möjligt i utvärderingsversionen av provet. Detta tillgodoses i det aktuella fallet av de ca 875 personer som deltog den 15 augusti.

Likvärdiga kravgränser mellan olika provversioner

En av de mest centrala och utmanande komponenterna i provutveckling är att fastställa gränser för olika provbetyg. I fallet med de olika delproven i Medborgarskapsprovet handlar det om en gräns, den för ett godkänt resultat. I styrdokumentet till vad provet ska omfatta



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

finns inga anvisningar om var en sådan gräns ska dras. Det finns inte heller några anvisningar om huruvida gränsen ska sättas i förväg eller om det bör göras efter provens genomförande. Vår bestämda uppfattning är att det är bättre att fastställa gränsen efteråt, eftersom ett mycket rikare datamaterial då finns tillgängligt.

Efter provgenomförandet analyseras hur uppgifterna har fungerat empiriskt. En expertgrupp bedömer därefter vilken prestation som motsvarar gränsen för godkänt. Bedömningarna vägs samman med information om uppgifternas svårighetsgrad och deltagarnas svarsmönster, och statistiska modeller kan användas som ytterligare stöd. Att gränsen fastställs efter genomförandet är särskilt viktigt eftersom olika provversioner sällan har exakt samma svårighetsgrad. Med hjälp av empiriska data och ankaruppgifter kan nya versioner länkas till tidigare prov, så att samma underliggande kravstandard kan upprätthållas även om råpoängsgränsen varierar. Nedan beskrivs processen för att fastställa gränsen för godkänt mer ingående.

Det är relativt vanligt, i kriterierelaterade betygssystem, att gränser för provbetyg meddelas i förväg, innan proven har genomförts. Så har till exempel skett i det svenska nationella provsystemet under många år. Skolverket har dock visat att andelen elevprestationer som motsvarar de olika provbetygen har varierat relativt mycket mellan olika versioner av samma prov, beroende på skolämne.⁴ Detta beror i huvudsak på att det är mycket svårt att i förväg uppskatta svårighetsgraden i en provversion när uppgifter i provet ändras mellan versioner. Om likvärdighet i svårighetsgrad över tid är avgörande, är rättssäkerheten lägre i ett system där betygsgränserna fastställs på förhand än om gränser sätt i efterhand. Det kan tyckas vara en fördel för en person som ska genomföra ett prov att veta kraven redan vid provtillfället, men det kan lika gärna utgöra en distraktion då en upplevd hög gräns kan utgöra ett stressmoment, medan en upplevd låg gräns kan leda till underskattning. En princip som bör användas vid provkonstruktion är att deltagarna ska ha så få distraktionsmoment som möjligt och i stället ges möjlighet att i så stor utsträckning som möjligt endast ägna sig åt att besvara frågorna.

⁴ <https://www.skolverket.se/getFile?file=3488>



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

När ett provtillfälle har genomförts undersöks inledningsvis om data tillåter att endast **en** gräns för godkänt resultat kan sättas. För detta krävs tillräcklig grad av endimensionalitet, att de i provet ingående provuppgifterna (egentligen svaren) är lokalt oberoende av varandra (om en person klarar en fråga ska det inte bero på att hen råkade kunna svaret på en annan fråga i samma prov) samt att andelen korrekta svar på varje item ökar med ökande totalpoäng på provet.⁵ När dessa krav är uppfyllda tar arbetet med att fastställa gränsen för godkänt resultat vid. Detta sker i ett samarbete med expertgrupper, inklusive erfarna lärare. Dessa får i uppdrag att med någon eller flera metoder, indirekt eller direkt, föreslå gränser.⁶ På senare tid har andra statistiska modeller utvecklats som kan ge stöd i arbetet med kravgränssättning. En sådan modell är 1-PLCDM, som i en endimensionell tillämpning kan användas för att skilja mellan två latenta klasser, exempelvis godkänt och icke godkänt. Modellen estimerar de parametrar som bäst förklarar de observerade svarsmönstren i den analyserade gruppen. Därefter kan varje individs provresultat klassificeras som tillhörande den latenta klass (godkänt respektive ej godkänt) som har högst sannolikhet givet individens svarsmönster och den estimerade modellen.⁷ Oavsett metoder är det dock i slutändan alltid ett beslut som tas av människor.

Det är sannolikt inte möjligt att använda samma version av ett delprov vid flera olika tillfällen, eftersom risken för spridning av innehållet är för stor. Delproven måste alltså bjudas i olika versioner. Nya versioner behöver ekvivaleras mot den första så att provtagarna, över tid, får så jämbördiga utmaningar som möjligt.⁸ Vår bedömning är att högst likvärdighet och

⁵ Wes Bonifay, *Unidimensional Item Response Theory* (SAGE Publications, Inc., 2020), 5–26, <https://doi.org/http://dx.doi.org/10.4135/9781506384276>.

⁶ Stephen G. Sireci och Alejandra Garcia, *Angoff Method* (SAGE Publications, Inc., 2018), 93–95, <https://doi.org/http://dx.doi.org/10.4135/9781506326139.n44>; Gregory J. Cizek m.fl., "Setting Performance Standards: Contemporary Methods", *Educational Measurement: Issues and Practice* 23, nr 4 (2004): 31–31, <https://doi.org/10.1111/j.1745-3992.2004.tb00166.x>; Jos Keuning m.fl., *The Data-Driven Direct Consensus (3DC) Procedure: A New Approach to Standard Setting*, red. Sigrid Blömeke och Jan-Eric Gustafsson (Springer International Publishing, 2017), 263–78, http://link.springer.com/10.1007/978-3-319-50856-6_15.

⁷ Matthew J. Madison m.fl., "A One-Parameter Diagnostic Classification Model with Familiar Measurement Properties", *Journal of Educational Measurement* 61, nr 3 (2024): 408–31, <https://doi.org/10.1111/jedm.12390>.

⁸ Joseph Ryan och Frank Brockman, *A Practitioner's Introduction to Equating* (Council of Chief State School Officers, 2009), <https://files.eric.ed.gov/fulltext/ED544690.pdf>.



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

rättssäkerhet i olika versioner av ett prov erhålls genom att även de nya gränserna sätts i efterhand. Argumentet för detta är även nu att det är först då maximal informationsmängd finns tillgänglig. För att underlätta arbetet finns statistiska metoder att tillgå, exempelvis frekvensbaserade metoder som bygger på att jämföra percentiler,⁹ parametriska metoder såsom *Stocking-Lord*¹⁰ och *Haebara*,¹¹ samt metoder baserade på falskt/sant positiva och negativa utfall.¹² Det är högst troligt att delproven kommer att besvaras av icke-ekvivalenta grupper. För att åstadkomma en ekvivalering är det därför nödvändigt med en viss andel gemensamma uppgifter – ankaruppgifter – i de olika versionerna av provet. En gemensam andel på minst 20 % men gärna uppemot 40 % brukar rekommenderas.¹³ En rimlig strategi kan vara att börja med en andel uppemot 40 % för att undersöka om skattningarna blir tillräckligt precisa för att därefter minska andelen efter hand. Sannolikt kommer också de gemensamma uppgifterna att spridas över tid och behöver bytas ut. Det kan göras successivt för att behålla en koppling till det ursprungliga.¹⁴

Sammanfattningsvis innebär detta att det mest likvärdiga tillvägagångssättet är att fastställa gränsen för godkänt resultat i efterhand. Råpoänggränsen kan variera mellan provversionerna, men den underliggande kravstandarden är avsedd att vara densamma. Enligt vår mening bör därför av kvalitetsskäl – likvärdighet, rättvisa och rättssäkerhet – standardsättningen av MBP ske enligt följande:

- Resultat från alla deltagares svar på alla enskilda uppgifter samlas in och analyseras.

⁹ Samuel A. Livingston, *Equating Test Scores (Without IRT) Second Edition* (2014), 73, <https://www.ets.org/Media/Research/pdf/LIVINGSTON2ed.pdf>.

¹⁰ Martha L. Stocking och Frederic M. Lord, "Developing a Common Metric in Item Response Theory", *Applied Psychological Measurement* 7, nr 2 (1983): 201–10, <https://doi.org/10.1177/014662168300700208>.

¹¹ Tomokazu Haebara, "Equating Logistic Ability Scales by a Weighted Least Squares Method", *Japanese Psychological Research* 22, nr 3 (1980): 144–49, <https://doi.org/10.4992/psycholres1954.22.144>.

¹² Maria Bolsinova och Gunter Maris, "Can IRT Solve the Missing Data Problem in Test Equating?", *Frontiers in Psychology* 6 (januari 2016), <https://doi.org/10.3389/fpsyg.2015.01956>.

¹³ Se t.ex. Davide Marengo m.fl., "Placing Multiple Tests on a Common Scale Using a Post-test Anchor Design: Effects of Item Position and Order on the Stability of Parameter Estimates", *Frontiers in Applied Mathematics and Statistics* 4 (oktober 2018), <https://doi.org/10.3389/fams.2018.00050>.

¹⁴ Erika Majoros, *Linking Recent and Older IEA Studies on Mathematics and Science* (2022), <https://hdl.handle.net/2077/71965>.



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

- Provets helhetsegenskaper analyseras.
- Standardsättning sker med hjälp av såväl statistiska modeller som expertgrupper.
- För efterföljande prov sätts gränsen efter provets genomförande med hjälp av länknings- och ekvivaleringsmetoder med utgångspunkt i gemensamma ankaruppgifter mellan provversioner samt tillsammans med expertbedömning.
- För kommande prov kan eventuellt resultat från tidigare provomgångar användas för att ange ett sannolikt intervall för kravgränsen.

Samhällskunskapsprovet specifikt

Tidplan, framtagande och genomförande

I rapport till UHR hösten 2025 presenterades en preliminär tidplan för utvecklingen av delprovet i samhällskunskap. Denna tidplan kvarstår i huvudsak oförändrad.

Utprovningstillfället den 15 augusti 2026 utgjorde den första utprovningen med den avsedda målgruppen. Under arbetsåret har uppgifter prövats i flera grupper, men av förklarliga skäl dock ännu inte i den specifika målgruppen för medborgarskapsprovet.

Under förutsättning att utprovningen i augusti 2026 ger tillfredsställande resultat bedömer provgruppen att förutsättningarna för att genomföra ett första skarpt delprov i mars 2027 stärks. Om ytterligare utvecklings- och utprovningsarbete bedöms nödvändigt, bör det första skarpa delprovet i stället genomföras i samband med de första delproven i svenska hösten 2027.

Provgruppen bedömer vidare att det på sikt bör vara möjligt att genomföra delprovet två gånger per år, förslagsvis i mars och september. Detta förutsätter att utprovningssuppgifter integreras i ordinarie provgenomförande för att möjliggöra kontinuerlig kvalitetssäkring och utveckling av uppgifter, likt genomförandet i augusti 2026.



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

När är digitalt prov möjligt?

Provgruppen genomför redan i dag en del av utprovningarna digitalt, eftersom detta underlättar insamling och analys av resultatdata.

Ett digitalt prov bedöms kunna införas tidigast två år efter det att UHR har upphandlat och implementerat en ändamålsenlig digital provplattform. Plattformen behöver genomgå omfattande tester och utprovningar för att säkerställa funktionalitet, rättssäkerhet och informationssäkerhet. Det är även nödvändigt att provgruppen ges möjlighet att testa plattformens funktionalitet i relation till provkonstruktion och genomförande.

Övergången från analoga till digitala prov innebär omfattande organisatoriska, tekniska och rättsliga överväganden. Det gäller bland annat hantering av personuppgifter, distribution av provmaterial, säker identifiering av provdeltagare, användning av låsta provmiljöer samt ansvarsfördelning mellan UHR och andra aktörer.

UHR behöver även ta ställning till om digitala enheter ska tillhandahållas centralt eller om provdeltagare ska kunna använda egna enheter. En modell där egna enheter används ställer höga krav på teknisk kompatibilitet, säkerhetslösningar och användarstöd.

Provgruppen bedömer att införandet av digitala prov kommer att kräva betydande resurser och ett långsiktigt utvecklingsarbete, framför allt hos den ansvariga genomförandemyndigheten.

När det gäller erfarenheter av digitala prov kan relevant kunskap hämtas från exempelvis Skolverket och Trafikverket.

Digitalt prov med uppgifter ur frågeförråd

Det är i nuläget inte möjligt att bedöma när ett digitalt prov baserat på ett förråd av uppgifter skulle kunna införas. Det är dock tänkbart att på sikt bygga upp ett förråd med uppgifter och data om dem. En praktisk lösning för ett sådant förråd skulle kunna vara tvådelad, där en del



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

består av själva uppgifterna och en annan del av metadata om dem. Det tar emellertid sannolikt flera år att kalibrera metadata så att de blir tillräckligt stabila. Viktigt att påpeka är också att valet ur ett sådant, eventuellt förråd måste ske av personer med stor metodologisk kompetens.

Ett sådant system skulle förutsätta omfattande och kontinuerliga utprovningar där varje enskild uppgift analyseras och kalibreras avseende exempelvis svårighetsgrad och funktionalitet. Vidare behöver, under ett antal inledande omgångar, populationen vara rimligt stabil för att skattningar ska bli hållbara över tid. Nedan följer ett tankeexempel:

Låt oss säga att ungefär 20 uppgifter kan prövas ut per provomgång och att det ges två prov per år. Vid 40 uppgifter per provomgång krävs det sannolikt minst 150 stabila uppgifter för att kunna göra ett urval som täcker hela domänen. Låt oss också säga att om 60 % av de utprovade uppgifterna överlever granskning och analys, tillförs cirka 12 uppgifter per provversion. Samtidigt kommer också uppgifter att behöva tas bort ur förrådet, beroende på att de antingen har exponerats för mycket eller åldrats av skäl som till exempel har att göra med att samhället förändras. Det är mycket svårt att förutse hur många uppgifter som kommer att behöva tas bort, men vi kan anta att det rör sig om ett par per provomgång. Sammantaget skulle det i så fall ta cirka 5 år att bygga upp ett förråd med tillräckligt många stabila uppgifter som skulle kunna användas för att välja uppgifter till nya provversioner med viss förutsägbarhet när det gäller svårighetsgrad. Till detta kommer att de första omgångarna av provtagare sannolikt kommer att skilja sig något från senare omgångar, vilket gör att förrådet kommer att behöva kalibreras på nytt efter ett antal år innan data om uppgifterna går att fastställa. Därutöver kommer de enskilda uppgifterna i varje provversion behöva kontrolleras så att de inte driver i väg när det gäller uppskattningarna av svårighetsgrad och diskriminering.

Om målsättningen är att på lång sikt kunna utveckla ett system med ett förråd av uppgifter, behöver detta tydliggöras i ett tidigt skede, eftersom det medför betydande behov av resurser, kompetensförsörjning och långsiktig förvaltning. Det handlar här inte minst om vem som står för förrådet samt vilken infrastruktur som behövs för meta-information och förvaltning. För



GÖTEBORGS UNIVERSITET INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

att få likvärdiga godkänt-gränser i olika provversioner behöver man högst sannolikt, även med ett förråd av uppgifter, fastställa slutgiltiga gränser efter det att proven har genomförts.

På ett övergripande plan är det viktigt att beakta att samhällskunskap är ett ämnesområde som påverkas av förändringar i lagstiftning, politik och samhällsutveckling. Detta har effekter såväl på utbildningsmaterialet som på proven. Uppgifter riskerar därför att bli inaktuella över tid, vilket innebär att det kontinuerligt kommer att finnas behov av utveckling och ersättning av uppgifter både i *Sverige i fokus* och i proven. Aspekter av spridning måste beaktas liksom risken för, och effekter av, kommersiella aktörers försök att – ev. med hjälp av AI – producera övningsmaterial för provet på basis av utbildningsmaterialet (se t.ex. <https://www.medborgarskaps-prov.se/ovningsprov>). Vidare begränsar materialets utformning möjligheterna till variation över tid, vilket innebär en stor risk för att materialet, och därmed provet, 'dör'. Detta understryker vikten av en snar granskning och vidareutveckling av utbildningsmaterialet *Sverige i fokus*.

Genomförande av delprovet samhällskunskap från och med 2027

Under förutsättning att utprovningen som genomfördes augusti 2026 ger tillfredsställande resultat bedömer provgruppen att delprovet kan genomföras två gånger per år, förslagsvis i mars och september.

Vi ser den framtida, reguljära provtiden som 60 minuter, med 20-30 min extra i de fall anpassning behövs. Delprovet bedöms omfatta cirka 40 uppgifter exklusive ett mindre antal utprovningssuppgifter. Utprovningssuppgifterna används för att pröva nya uppgifter inför framtida prov och påverkar inte provdeltagarens resultat.

Kostnadsberäkning

Kostnadsberäkningen baseras på nuvarande personalstyrka och nuvarande organisation av arbetet.



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Under uppbyggnadsfasen bedöms behovet av personella resurser vara relativt omfattande. Det är samtidigt svårt att i nuläget göra långsiktiga bedömningar av framtida resursbehov och kostnader, eftersom flera avgörande ställningstaganden ännu inte har gjorts av UHR.

Det gäller exempelvis omfattningen av framtida digitalisering, val av tekniska lösningar samt ansvarsfördelning mellan olika aktörer. Dessa faktorer påverkar såväl utvecklingskostnader som framtida drift och förvaltning.

En annan faktor som kan påverka kostnader är hur ofta och i vilket skede som nödvändig uppdatering av materialet *Sverige i fokus* äger rum. Materialet ska ligga till grund för de kunskaper som testas i provet. Om materialet ändras kontinuerligt och tätt inpå när prov ska genomföras så påverkar detta både processen kring uppgiftskonstruktion samt processen när ett prov sätts samman, vilket kommer att påverka kostnaderna.

I nuläget uppskattas kostnaderna för att ta fram delprov i samhällskunskap till cirka 4 miljoner per år. Från hösten 2026 kommer drygt två heltidstjänster att omfattas av uppdraget; behovet kan behöva ökas under en tid på grund av implementering av digitala delprov.

Vid förändring av samordningsuppdraget måste vissa funktioner, exempelvis GERS-funktion och metodstöd ökas på i gruppen, vilket kommer att medföra ökade kostnader. Detta kommer att vid behov kommuniceras med UHR i god tid.

Referenser

Bolsinova, Maria, och Gunter Maris. "Can IRT Solve the Missing Data Problem in Test Equating?" *Frontiers in Psychology* 6 (januari 2016).
<https://doi.org/10.3389/fpsyg.2015.01956>.

Bonifay, Wes. *Unidimensional Item Response Theory*. SAGE Publications, Inc., 2020.
<https://doi.org/http://dx.doi.org/10.4135/9781506384276>.



GÖTEBORGS UNIVERSITET
INST FÖR PEDAGOGIK OCH SPECIALPEDAGOGIK

Cizek, Gregory J., Michael B. Bunch, och Heather Koons. "Setting Performance Standards: Contemporary Methods". *Educational Measurement: Issues and Practice* 23, nr 4 (2004): 31–31. <https://doi.org/10.1111/j.1745-3992.2004.tb00166.x>.

DeMars, Christine, 'Requirements', Item Response Theory (2010; online edn, Oxford Academic, 22 Mar. 2012), <https://doi.org/10.1093/acprof:oso/9780195377033.003.0002>

Haebara, Tomokazu. "Equating Logistic Ability Scales by a Weighted Least Squares Method". *Japanese Psychological Research* 22, nr 3 (1980): 144–49. <https://doi.org/10.4992/psycholres1954.22.144>.

Keuning, Jos, J. Hendrik Straat, och Remco C. W. Feskens. *The Data-Driven Direct Consensus (3DC) Procedure: A New Approach to Standard Setting*. Redigerad av Sigrd Blömeke och Jan-Eric Gustafsson. Springer International Publishing, 2017. http://link.springer.com/10.1007/978-3-319-50856-6_15.

Livingston, Samuel A. *Equating Test Scores (Without IRT) Second Edition*. 2014. <https://www.ets.org/Media/Research/pdf/LIVINGSTON2ed.pdf>.

Madison, Matthew J., Stefanie A. Wind, Lientje Maas, Kazuhiro Yamaguchi, och Sergio Haab. "A One-Parameter Diagnostic Classification Model with Familiar Measurement Properties". *Journal of Educational Measurement* 61, nr 3 (2024): 408–31. <https://doi.org/10.1111/jedm.12390>.

Majoros, Erika. *Linking Recent and Older IEA Studies on Mathematics and Science*. 2022. <https://hdl.handle.net/2077/71965>.

Marengo, Davide, Renato Miceli, Rosalba Rosato, och Michele Settanni. "Placing Multiple Tests on a Common Scale Using a Post-test Anchor Design: Effects of Item Position and Order on the Stability of Parameter Estimates". *Frontiers in Applied Mathematics and Statistics* 4 (oktober 2018). <https://doi.org/10.3389/fams.2018.00050>.

Ryan, Joseph, och Frank Brockman. *A Practitioner's Introduction to Equating*. Council of Chief State School Officers, 2009. <https://files.eric.ed.gov/fulltext/ED544690.pdf>.

Sireci, Stephen G., och Alejandra Garcia. *Angoff Method*. SAGE Publications, Inc., 2018. <https://doi.org/http://dx.doi.org/10.4135/9781506326139.n44>.

Stocking, Martha L., och Frederic M. Lord. "Developing a Common Metric in Item Response Theory". *Applied Psychological Measurement* 7, nr 2 (1983): 201–10. <https://doi.org/10.1177/014662168300700208>