



GÖTEBORGS
UNIVERSITET

Ett gränsfall?

En undersökande studie om konsekvenserna av en icke-kompensatorisk modell för kravgränssättning på de nationella proven i matematik 2c.

Namn: Carlos Abarca och Jenny Arkevall
Program: Kompletterande Pedagogisk Utbildning



Uppsats/Examensarbete: 15 hp
Kurs: LKXA1G
Nivå: Grundnivå
Termin/år: VT / 2016
Handledare: Peter Nyström
Examinator: Rimma Nyman
Kod: VT16-2930-LKXA1G-001

Nyckelord: Kravgränssättning, nationella prov, matematik, icke-kompensatorisk, kompensatorisk, provbetyg, kursbetyg.

Abstract

De nationella proven i matematik har idag en kompensatorisk kravgränssättning i förhållande till de förmågor som enligt läroplanen skall bedömas icke-kompensatoriskt då kursbetyg bestäms. Denna undersökning behandlar ämnet om huruvida det är möjligt att applicera en icke-kompensatorisk modell för betygsättning på dagens nationella prov i matematik.

Vi har analyserat två nationella prov i matematik 2c på gymnasiet och har utifrån dessa räknat ut nya kravgränser med två icke-kompensatoriska metoder. De resultat vi kommit fram till är att det är fullt möjligt att ha en kravgränssättning som är icke-kompensatorisk på detta vis, dock skulle detta betyda ett sänkt betygssnitt på provet, trots att kravgränserna vid en snabb anblick kan se enklare ut.

1 Förord

Som lärarstudenter har vi lärt oss att det inte är lätt att sätta betyg. När vi under bedömningskursen började studera de nationella proven insåg vi att det finns en motsättning mellan hur kursbetyg och provbetyg sätts, vilket resulterade i många långa diskussioner. Vi lyckades dock inte nysta upp grunden till problemen och inga frågor resulterade i entydiga svar. Vårt fria fall, som var arbets gång, utför bedömningsstupet resulterade i slutändan dock i flera insikter som inte hade besvarats i en vardaglig situation i yrket, vilket vi är glada för.

Vi vill framför allt tacka Anna Lind Pantzare, projektledare för de nationella proven på Umeå Universitet, samt vår handledare Peter Nyström för er hjälpsamhet, snabba svar och ljuslykta när vi famlade i mörkret.

2 Innehållsförteckning

INLEDNING	1
SYFTE OCH FRÅGESTÄLLNINGAR	2
BEGREPP	3
TEORETISK BAKGRUND	5
KRAVGRÄNSSÄTTNING - <i>STANDARD SETTING</i>	5
<i>Kravgränsmodeller</i>	5
<i>Kravgränssättningsmetoder</i>	6
RELIABILITET I KUNSKAPSMÄTNINGAR.....	9
METOD	10
GENOMFÖRANDE	10
VAL AV KRAVGRÄNSSÄTTNINGSMETODER	11
<i>Vår modifierade Angoff-metod</i>	11
<i>Vår modifierade Contrasting Groups-metod</i>	12
RESULTAT	14
DISKUSSION	16
RESULTATDISKUSSION.....	16
<i>Svårighetsgraden i poängfördelningen</i>	16
<i>Totalt antal poäng</i>	17
<i>Kompensatoriska kontra icke-kompensatoriska kravgränser</i>	17
METODDISKUSSION	18
SLUTSATSER OCH VIDARE FORSKNING	19
REFERENSLISTA	20
BILAGA 1 – POÄNGFÖRDELNING VT-15	
BILAGA 2 – DIAGRAM FÖR KRAVGRÄNSER VT-14	
BILAGA 3 – DIAGRAM FÖR KRAVGRÄNSER VT-15	

3 Inledning

I läroplanen för gymnasieskolan står det under matematikämnets syfte att undervisningen ska ge eleverna förutsättning att utveckla de förmågor som benämns *begrepps-, procedurs-, problemlösnings-, modellering-, resonemang-, kommunikations- och relevansförmåga* (Skolverket, 2011; Skolverket, 2016a). Förmågorna uttrycks på olika nivåer och ligger till grund för de kunskapskrav som formulerats för betygen E, C och A. Dessa kunskapskrav preciserar enligt gymnasieförordningen de kunskaper som krävs för ett visst betyg. Skolverket (2012) anger att **alla** kunskapskrav för en viss nivå måste vara uppfyllda för att detta betyg skall kunna nås¹. Kursbetygen är således icke-kompensatoriska, då brister i en förmåga inte kan kompenseras med styrkor i andra (se kapitel 6.1.1 för en närmare beskrivning av begreppet).

De nationella proven i matematik följer idag en kompensatorisk modell med avseende på förmågorna när betyg sätts på provet. I bedömningsanvisningar anges i en matris att varje uppgift är markerad med den förmåga poängen mäter. Dock tillkommer i slutskedet ett provbetyg, där varje poängfördelning efter förmåga inte tas hänsyn till (Umeå Universitet, 2016). I det avseendet är modellen för provbetyget kompensatorisk, då inga krav på kunskaper per förmåga krävs.

På Skolverkets (2016b) webbsida behandlas frågan om *relationen mellan provbetyg och betyg* och varför dessa inte är jämförbara. Där beskrivs att de nationella proven huvudsakligen uppfyller **två** syften: att ”stödja en likvärdig och rättvis bedömning och betygsättning” samt att ”ge underlag för en analys av i vilken utsträckning kunskapskraven uppfylls på skolnivå, på huvudmannanivå och på nationell nivå.” Vidare anges att betygen ”används för att möjliggöra de jämförelser som efterfrågas på lokal och nationell nivå. De provbetyg som ingår i proven, och ibland i delproven, får inte förväxlas med det betyg som sätts i slutet av en termin eller kurs” (Skolverket, 2016b). Dessa två syften är dock oförenliga enligt Statens officiella utredning (SOU, 2016:25) som föreslår att proven endast skall ha **ett** huvudsyfte.

Wikström (2013) hävdar att ”Det allra viktigaste är att se till att syftet med provet harmoniserar med dess konstruktion och användning.” (s. 172), och lyfter fram tre frågor som är angelägna att besvara. Först berörs det huvudsakliga syftet med provet samt huruvida konstruktionen passar med syftet. Vidare bör frågan om huruvida provet ger den information som behövs besvaras. Hon menar att om syftet med provet inte är tydligt, eller om de två senare frågorna inte säkert kan besvaras med ett *ja*, bör provet omarbetas (Wikström, 2013).

Om förslaget från den nyligen publicerade utredningen om nationella prov som nämnts ovan (SOU, 2016:25) skulle antas, läggs frågan om att sätta upp ett ramverk för de nationella proven över till en av Skolverket tillsatt grupp. En av frågorna denna grupp bör ta ställning till är hur provbetygen för de nationella proven skall vara utformade – om de överhuvudtaget ska finnas kvar (SOU, 2016:25).

¹ För betygen D och B krävs att alla kriterier för E respektive C, samt övervägande del av C och A, skall uppfyllas.

4 Syfte och frågeställningar

Med utgångspunkt i problembilden som beskrivs ovan kommer detta examensarbete att behandla hur en icke-kompensatorisk betygsättning på de nationella proven i matematik skulle kunna fungera. Syftet med uppsatsen är att testa huruvida en sådan kravgränsmodell kan appliceras på det nationella provet som det ser ut idag.

Följande frågeställningar kommer således utforskas:

1. Är det möjligt att applicera en icke-kompensatorisk kravgränssättningsmetod på de nationella proven i matematik, så som de är formade idag?
2. Vad skulle hända om en sådan kravgränssättningsmetod appliceras?
3. Hur skulle kravgränserna i sådana fall se ut?

5 Begrepp

Några begrepp används genomgående i arbetet och är viktiga att särskilja. Viktigt att betona för läsaren är att nedanstående är en kort beskrivning av begrepp, en utförligare definition förekommer i vissa fall senare i arbetet. Detta är en samlad kunskap som vi har dragit från Cizek och Bunch (2007), Crocker och Algina (2008), Mehrens och Cizek (2012), Haladyna och Hess (1999) samt Hambleton (1995).

- **Kravgränssättningsmetod** - Syftar till metoder att sätta kravgränser i form av en poängsumma. Kategoriseras som *analytiska* och *holistiska*.
 - **Analytiska** - Fokuserar på prestation **per uppgift**. En distinktion görs mellan *test-centrerade* eller *provtagar-centrerade*.
 - **Test-centrerade** - Utgår från provet. Gör uppskattningar på hur en tänkt elev skulle prestera. Exempelvis: Angoffs metod
 - **Provtagar-centrerade** - Utgår från provtagarnas faktiska prestation. Vi anser att vår modifierade Angoff-metod är provtagar-centrerad.
 - **Holistiska** - Fokuserar kring prestation på en totalpoäng **hela provet**.
 - **Test-centrerade** - Utgår från provet. Gör uppskattningar över hur en tänkt elev skulle prestera.
 - **Provtagar-centrerade** - Utgår från provtagarnas faktiska prestation. Ex: Contrasting Groups-metoden.
- **Kravgränsmodell**² - Syftar till olika sätt att använda kravgränssättningsmetoder för en betygsättning. De modeller som finns är *icke-kompensatorisk* respektive *kompensatorisk modell*.
 - **Kompensatorisk modell** – Bristande kunskap inom ett delområde kan kompenseras med god kunskap inom ett annat. Kravgräns är satt på totala antalet poäng.
 - **Icke-kompensatorisk modell** – Kunskap inom specifika delområden på provet krävs. Bristande kunskap inom detta kan inte kompenseras med goda kunskaper inom ett annat delområde.
 - **Konjunktiv** - Hänsyn till alla delområden tas. Godkänt på alla dessa krävs.

² Begreppet används få gånger i uppsatsen. Framst har begreppet formulerats för att kunna särskilja det från *kravgränssättningsmetoder*, vilket underlättar när olika modeller kombineras med metoder.

- **Disjunktiv** – Vissa delområden eller en delmängd av kritiska aspekter måste avklaras för godkänt.
- **Kriteriekategori** – Den kombination av förmågorna som gjorts utifrån de fyra styckena i kunskapskraven. Tre kriteriekategorier har distingerats. *Begrepp/Procedur, Problemlösning/Modellering* samt *Kommunikation/Resonemang*.

6 Teoretisk bakgrund

För att kunna besvara forskningsfrågorna krävs kunskap kring vad en kravgränssättning innebär samt vilka metoder och modeller det finns för att bestämma kravgränser på prov. Avslutningsvis tar vi i detta kapitel upp frågor som allmänt rör reliabilitet och då främst vad gäller förändringar utav den på prov. Detta är något som framförallt återkommer i en diskussion kring de valda metoderna.

6.1 Kravgränssättning - *Standard setting*

Det som på engelska kallas *standard setting* är vetenskapen om hur kravgränser sätts på prov eller test (Crocker & Algina, 2008, s. 412). Enligt Cizek och Bunch (2007) kan kravgränssättning definieras som “a procedure that enables participants using a specified method to bring to bear their judgments in such a way as to translate the policy positions of authorizing entities into locations on a score scale” (s. 19). Kravgränsmodeller syftar till användningen av metoder för att beräkna kravgränser lämpade efter vald modell.

6.1.1 Kravgränsmodeller

Det finns två olika kravgränsmodeller vilka benämns som *icke-kompensatorisk* respektive *kompensatorisk modell*. Dessa modeller skiljer sig åt genom att de tar hänsyn till kritiska aspekter på olika sätt, förutsatt att det finns några att ta hänsyn till. Med kritiska aspekter menas – i samband med kravgränsmodeller – helt enkelt något som gör att en uppgift kan skiljas från en annan. Det kan vara att uppgifter behandlar en specifik förmåga, är av en viss svårighetsgrad eller behandlar ett specifikt centralt innehåll.

I en kompensatorisk modell för kravgränssättning tas det **inte** hänsyn till dessa kritiska aspekter. Trots att uppgifterna kan behandla olika kunskaper, förmågor eller innehåll hanteras alla uppgifter likvärdigt. Kravgränser bestäms enligt en kravgränsmetod baserad på alla uppgifter. I en icke-kompensatorisk modell för kravgränssättningen tas det däremot hänsyn till de kritiska aspekterna. Detta kan göras *konjunktivt* eller *disjunktivt*, där konjunktivt menas att en uppfyllelse inom alla aspektområden tas i beaktande och med disjunktivt menas att endast en delmängd av dem beaktas (Cizek & Bunch, 2007; Crocker & Algina, 2008; Haladyna & Hess, 1999; Hambleton, 1995; Mehrens & Cizek, 2012).

Det nationella provet i matematik är intressant fall att titta på i samband med detta, då uppgifterna på provet redan är indelade i kategorier som förmågor och svårighetsnivå. Med avseende på förmågorna följer det nationella provet en kompensatorisk modell, då det inte ställs några krav på antal poäng inom specifika förmågor för att få ett visst betyg. Däremot ställs det krav för vissa betygssteg att en delmängd av uppgifter inom en viss svårighetsnivå ska vara uppfyllda. I det avseendet följer det nationella provet en icke-kompensatorisk modell, mer specifikt en disjunktiv icke-kompensatorisk modell.

Här följer en illustration av kompensatorisk kontra icke-kompensatoriska modeller. A, B och C är poäng en provtagare får inom tre olika aspekter på ett prov. För ett visst betyg måste provtagaren enligt modellen uppfylla:

Kompensatorisk:

Summan av (A, B och C) \geq X, där X är kravgränsen för det avsedda betyget.

Icke-kompensatorisk (konjunktiv):

(A \geq x) **och** (B \geq y) **och** (C \geq z), där x, y och z är kravgränser inom de olika aspekterna för det avsedda betyget.

Icke-kompensatorisk (disjunktivt):

(A \geq x) **eller** (B \geq y) **eller** (C \geq z), där x, y och z är kravgränser inom de olika aspekterna för det avsedda betyget.

Haladyna och Hess (1999) har studerat jämförelser mellan kompensatoriska och konjunktivt icke-kompensatoriska modeller³. De konstaterar att reliabiliteten i en konjunktiv modell aldrig kan överstiga reliabiliteten i kompensatorisk modell. Detta är logiskt då i en konjunktiv modell är varje område bedömd på mindre information än helheten. Vidare konstaterar de att en konjunktiv modell kommer ställa högre krav än en kompensatorisk modell samtidigt som den kompensatoriska modellen inte ger någon diagnostisk information om elevens utvecklingspotential. De hävdar även att det i en konjunktiv modell varken är användbart eller meningsfullt att ha en totalpoäng eftersom alla områden är nödvändiga och lika viktiga. (Haladyna & Hess, 1999, ss. 135-136).

6.1.2 Kravgränssättningsmetoder

Enligt Crocker och Algina (2008) finns över 30 moderna metoder för att beräkna kravgränser, vilka dock alla faller under tre huvudkategorier: *analytiska*, *holistiska* samt *provtagarcentrerade metoder*. Som beskrivits ovan kan analytiska och holistiska metoder klassas både som provtagarcentrerade eller testcentrerade. De flesta metoderna är dock testcentrerade, varför provtagarcentrerade har fått en egen huvudkategori.

Analytiska metoder

Av de tre ovan nämnda kategorierna är analytiska metoder de mest studerade. Metoderna bygger på att en expertgrupp med god kunskap inom både ämnesområdet samt provtagarnas nivå tillsätts. Denna expertgrupp får sedan i uppgift att tänka sig en elev – eller grupp av elever – som de bedömer **precis** uppfyller kraven för godkänt (Crocker & Algina, 2008; Hambleton & Pitoniak, 2006). Expertgruppen skall enligt Cizek och Bunch (2007) vara en grupp med kvalificerad kunskap i både ämnet och innehållet och de skall helst göra testet själva i förväg.

³ Haladyna & Hess (1999) differentierar även mellan sekventiell och icke-sekventiell konjunktiv modell och det som refereras här gäller icke-sekventiell konjunktiv modell, men kan konstateras allmänt för konjunktiva modeller. För vidare läsning hänvisas den intresserade till deras studie.

De största nackdelarna med metoderna, förutom att det krävs många personer för att kunna bilda en expertgrupp, är tidsåtgången att genomföra metoderna (Crocker & Algina, 2008).

Av de analytiska metoderna är Angoffs metod vanligt förekommande, vilket även många andra analytiska metoder bygger på, eller är varianter av. Metoden bygger på Nedelskys metod, men har fått mer spridning än denna då den är applicerbar på fler typer av prov, då Nedelskys metod endast är applicerbar på prov med flervalsfrågor (Cizek & Bunch, 2007).

Holistiska metoder

De holistiska metoderna kräver däremot att en expert gör en helhetsbedömning av vad en provtagare behöver kunna för att ha klarat av provet. Det kan även göras med en grupp experter, och ju fler experter desto stabilare kravgränssättning. Experterna sätter en viss poäng för godkänd-nivån. Antigen är denna poängsättning dikotom, det vill säga godkänd/inte godkänd, men den kan även vara en rangordnad poängsättning, det vill säga implementeras då betyg sätts i en skala, exempelvis E–A (Cizek & Bunch, 2007; Crocker & Algina, 2008). Om en grupp experter gör bedömningarna, bestäms kravgränsen genom medelvärdet av deras bedömningar. Kritik har riktats mot osäkerheten i expertgruppernas bedömningar samt huruvida interbedömarreliabiliteten hos expertgruppen kan säkerställas. Då holistiska metoder väljs är det viktigt att noga dokumentera vilka experterna är samt deras kunskapsnivå, utbildning och yrke (Crocker & Algina, 2008).

Provtagarcentrerade metoder

De flesta kravgränssättningsmetoder – såväl analytiska som holistiska – kräver en expertgrupp för att göra bedömningar av diverse slag. Dock kan de sägas vara fokuserade kring ett prov och dess tänkta provtagare, därmed sägs vara test-centrerade. Det finns metoder som istället fokuserar på provtagaren och de benämns vara provtagar-centrerade. De går ut på att eleverna kategoriseras i grupper innan provet utförs, se nedan, och utifrån deras resultat bestäms en kravgräns.

Förespråkarna för dessa metoder argumenterar för att expertgruppen i fråga finner det mer välbekant med en provtagar-centrerad metod än en test-centrerad metod. Detta för att sådana expertgrupper – exempelvis lärare, läkare, professorer etc. – är vana vid att göra bedömningar på om en student besitter tillräcklig kunskap för att vara godkänd, eller annan gradering som man är ute efter (Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006).

Kritiker till metoderna påpekar faktorer såsom att provet måste genomföras innan en kravgräns kan sättas. Vidare belyser de svårigheterna för en expertgrupp att avgöra vilken nivå eleverna ligger på. Det är inte alltid tydligt vilken grupp provtagarna tillhör. Det finns även en problematik i att kravgränsen inte nödvändigtvis går att generalisera till en annan grupp provtagare än den som använts för att beräkna kravgränsen. Det har även riktats kritik mot hur expertgruppen utses samt deras utsatta position. De skall ha tillräcklig kunskap om provtagarnas kunskapsnivå, vilket enligt kritiken anses som en omöjlighet. Vidare finns en slags rundgångsargumentation i och med att provtagarnas förutbestämda kunskapsnivå bestämmer

kravgränsen, samtidigt som kravgränsen indikerar utifall provtagarna har tillräckligt med kunskap eller inte (Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006).

Som nämnts ovan finns många kravgränssättningsmetoder och ännu fler varianter på varje kravgränssättningsmetod. För undersökningens skull har endast ett fåtal valts att studeras och endast två kommer här att beskrivas. Nedan följer en beskrivning av de vanligaste varianterna av metoderna.

Angoffs metod

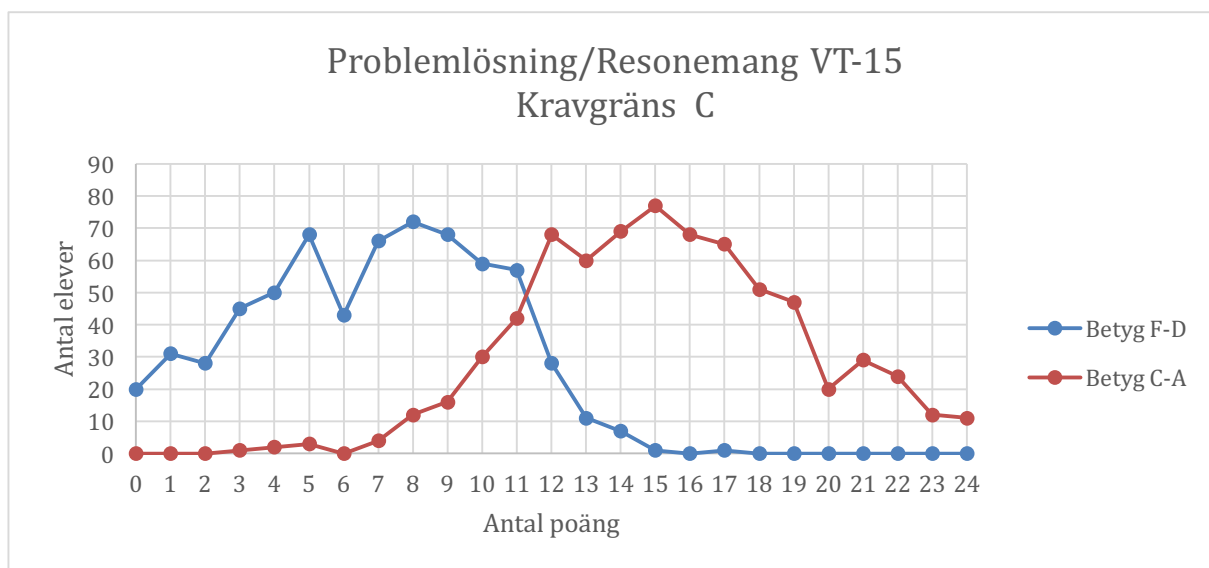
Angoffs metod bygger på Nedelskys men är mer generell i avseendet att den går att applicera på test/prov som inte bara omfattar flervalsfrågor. Metoden går ut på att en grupp bedömare tillsätts. Dessa tänker sig en provtagare som har precis acceptabel kunskapsnivå och bedömer **varje uppgift** utefter hur stor sannolikhet det är att en sådan person får full poäng (varje uppgift är värd ett poäng). Varje uppgift får således ett sannolikhetsvärde – uttryckt mellan 0 och 1 – vilka därefter summeras och bildar varje bedömares egen kravgräns. Medelvärdet (eller medianen) av dessa uppskattade kravgränser beräknas och sätts som lämplig kravgräns. En vanligare variant av Angoff-metoden går ut på att 100 provtagare, som har samma kunskapsnivå som ovan beskrivna provtagare, tänks. Expertgruppen bedömer hur många av dessa tänkta provtagare som skulle klara uppgiften, de skattade värdena läggs sedan ihop och avrundas till ett heltal, vilket motsvarar en lämplig kravgräns. Hur avrundningen bör ske är upp till bedömargruppen att avgöra, utifrån en gemensamt bestämd policy (Hambleton & Pitoniak, Setting performance standards, 2006; Cizek & Bunch, 2007; Livingston & Zieky, 1982).

Kritik mot metoden har främst handlat om att bedömare har en tendens att överskatta provtagarnas förmågor i enklare uppgifter medan de underskattar förmågan i de svårare. Vidare kritiseras hela metoden då det är en svår kognitiv uppgift att klara av att tänka både procentuellt hur många som klarar uppgiften i fråga, men även svårigheter att kunna tänka sig en person som ligger precis på denna gräns (Hambleton & Pitoniak, 2006). Om misstanke att expertgruppen har dessa svårigheter bör metoden, enligt Livingston och Zieky (1982) inte användas.

Contrasting groups metoden

Contrasting groups metoden, kräver en väl insatt expertgrupp kring både provtagarnas kunskapsnivå och materialet som bedöms. Expertgruppen gör en bedömning av provtagarnas nivå och kategoriserar dem i två grupper, en som bedöms ha tillräcklig kunskap och en som inte har det. Provtagarna tillhör således en av två grupper, F- eller E-grupp. Provtagarna utför sedan provet och utifrån **deras totala resultat** plottas två kurvor som visar antal poäng mot antal provtagare. Kravgränsen är den poäng som **lika många** ur respektive grupp fått, det vill säga den punkt där de två kurvorna korsar varandra (se Figur 1). I den poängen är det 50% chans att en elev tillhör E- eller F-gruppen (Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006). Då informationen innehåller faktiska resultat och elevprestationer är det inte säkert att de två kurvorna inte skär varandra mer än en gång vilket orsakar problematik. Ifall detta är fallet krävs metoder för regression, vilket vi inte går in närmare på här (Cizek & Bunch, 2007).

Figur 1- Exempel på en Contrasting Groups plottning.



6.2 Reliabilitet i kunskapsmätningar

Vid kunskapsmätningar måste vissa begränsningar tas i beaktning och enligt testteorin anses det observerade värdet, poängresultatet X_{pf} , vara lika med det sanna resultatet τ_p , plus ett fel E_{pf} . Heartel (2006) betecknar detta som att, om provtagare p gör prov X under administrativ form⁴ f , är $X_{pf} = \tau_p + E_{pf}$. Dessa fel kan exempelvis bero på vilken tid provet utförs eller hur provtagarna uppfattar provet. Felen kan även uppkomma i fråga om validitet på provet samt huruvida poängen speglar de olika kunskapsnivåer som bedöms.

För att kunna uppskatta sådana fel kan reliabiliteten mätas och en reliabilitetskoefficient kan beräknas, vilket dock inte kommer behandlas i denna uppsats. Det bör dock nämnas att denna koefficient betecknas ρ och är ett sannolikhetsvärde mellan 0 och 1. Enligt *Spearman-Browns prophecy formula* studeras reliabiliteten då längden på ett prov förändras, vilket kommer behandlas i uppsatsen. Formeln nedan beskriver hur stor den nya reliabiliteten ρ_{zz} är då provet förändras med faktor k , där $k < 1$ representerar en förkortning av provet och $k > 1$, en förlängning.

$$\rho_{zz} = \frac{k\rho_{xx}}{1 + (1 - k)\rho_{xx}}$$

ρ_{xx} är reliabiliteten innan längden på provet förändrats. Tydligt syns här att den nya reliabiliteten snabbt sjunker ju mindre k är, det vill säga om antalet uppgifter minskar, samt höjs om antalet uppgifter ökar, om k blir större (Heartel, 2006).

⁴ Med administrativ form menas alla yttre omständigheter kring provet.

7 Metod

För att besvara vår fråga har resultat från två nationella prov i matematik 2c utförda VT15, respektive VT14, undersökts. Dessa resultat användas för att simulera modifierade versioner av Angoffs metod och Contrasting-Groups metoden.

7.1 Genomförande

Data från Umeå universitet innehållande detaljerade information kring 2235 elevers resultat på det nationella provet i matematik 2c under vårterminen 2015, samt 1995 elever som skrev provet VT14, har använts för att på nya sätt räkna ut kravgränser på proven. Dessa data innehåller inte alla provtagares resultat det aktuella året, utan är slumpmässigt som statistiskt underlag vid en närmare utvärdering av provet. Information kring exakt vilka uppgifter varje provtagare har fått poäng på samt deras provbetyg och, på flertalet provtagare, även ett preliminärt kursbetyg finns i datamängden. Uppgifter med möjlighet till fler än en poäng delas in i ”deluppgifter” som ger en poäng vardera (se Bilaga 1). Vidare hittas information om kön samt om eleven har annat modersmål än svenska, detta är dock uppgifter som inte använts i vår undersökning. Datamängden för provet som gjordes till VT14 innehöll utöver ovanstående även information som elevers namn eller personnummer, vilket inte heller är av intresse, varför dessa av etiska skäl raderades. Datamängden har även kompletterats med information kring vilken kriteriekategori varje uppgift mäter.

I bedömningsanvisningarna anges för varje uppgift vilken slags förmåga som prövas (Umeå Universitet, 2016). De betonar dock att förmågorna inte är oberoende av varandra och att ”det är den förmåga som bedöms som den huvudsakliga” (2016, s. 15) som är den förmåga som uppgiften prövar. Kunskapskraven är däremot inte explicit uttryckta i förmågor, utan implicit kan de tidigare beskrivna förmågorna utläsas. Betygskriterierna är uttryckta på ett sådant sätt att det ibland är svårt att urskilja om ett specifikt kriterium handlar om en förmåga eller en annan. Dock är de formade på ett sådant sätt att betygskriterierna för E, C respektive A är formulerade i fyra stycken. Varje stycke är gradvis jämförbara – det vill säga att första stycket i betygskriterierna för E är jämförbara med kriterierna för C respektive A – och behandlar två förmågor. Första stycket handlar, enligt vår tolkning, om förmågorna begrepp och procedurer, andra om problemlösning och modellering, tredje om resonemang och kommunikation samt sista stycket behandlar hur matematiken relateras till verkligheten samt till viss del resonemang kring relevans.

Med bakgrund av det som framkommer i *Reliabilitet i kunskapsmätningar*, beslöts att inte applicera kravgränser på alltför många underkategorier, varför det valdes att kunskapskraven delas in i tre kriteriekategorier. Dessa kriteriekategorier valdes utifrån ovan beskrivna stycken och innehåller således förmågorna *begrepp-procedur*, *problemlösning-modellering* och *resonemang-kommunikation*. Vi har använt oss av den bedömning kring vilken förmåga respektive uppgift mäter, som idag finns på provet då det för arbetets ändamål inte finns någon anledning att ifrågasätta denna panels expertis.

Undersökningen går ut på att studera hur en icke-kompensatorisk⁵ kravgränsmodell står sig i förhållande till en kompensatorisk, varför båda dessa modeller används per kravgränssättningsmetod. Två kravgränser satta efter samma villkor, elevers faktiska resultat, erhålles då. Anledningen till att två olika kravgränssättningsmetoder används är att paralleller mellan jämförelserna på de kompensatoriska och icke-kompensatoriska kravgränssättningsmetoderna kan dras och således kan slutsatserna göras mer generella. Vidare ökas reliabiliteten i undersökningen av att två nationella prov undersöks då aspekter som provets konstruktion eller eventuella felmarginaler i datainsamlingen inte påverkar resultat lika mycket.

7.2 Val av kravgränssättningsmetoder

Efter att närmare ha studerat ett antal olika kravgränssättningsmetoder föll valet på att använda en modifierad variant på Angoffs kravgränssättningsmetod. Delvis valdes denna metod att modifieras, för att det är den som idag används då kravgränser bestäms på det nationella provet. Delvis valdes metoden för att den på ett naturligt sätt kan göras om till provtagarcentrerad och då utgå från faktiska resultat istället för en expertgrupps omdömen.

Det andra valet föll på används Contrasting Groups-metoden då denna metod kräver kunskap om elevers betygsnivå för att göra indelningar i grupper, vilket fanns i den erhållna datamängden i form av preliminär kursbetyg. De kontrasterande grupperna består således av elever som har fått preliminärt betyg F kontra E till A, för betygsgränsen E⁶.

7.2.1 Vår modifierade Angoff-metod

Då faktisk elevprestation – provresultat – finns tillgänglig kommer detta att användas istället för en expertgrupps skattningar av elevprestationer. De kravgränser som är satta på nationella provet som det ser ut idag, är framtagna av en expertgrupp. Denna grupps expertis ifrågasätts inte, faktum är att denna expertis är något som används i vår modifierade version av Angoffs metod. En elev som nått och jämt uppnått ett betyg anses nämligen inneha minimal accepterad kunskapsnivå för detta betyg. Sannolikheten att en elev, som precis uppnått kravgränsen, klarar av varje uppgift ur varje kriteriekategori beräknas och är således faktisk.

För att finna kravgränserna för E, C och A väljs de eleverna vars provresultat exakt uppfyller kraven respektive betyg enligt de satta kravgränserna. För kravgränsen E fokuseras de elever som fått totalt 14 poäng på provet, för kravgränsen C gäller 28 poäng varav minst 11 poäng på minst C-nivå. För kravgräns A är det på motsvarande sätt de elever som fått totalt 44 poäng varav minst 9 poäng på minst A-nivå, detta gäller för VT-15. Provet VT-14 fokuserar elever med poängen 14, 30 med 11 C poäng samt 45 varav 9 A poäng (Umeå Universitet, 2016).

⁵ Med icke-kompensatorisk avses i denna undersökning konjunktivt icke-kompensatorisk.

⁶ För kravgränsen för betyg C används elever som fått de preliminära betygen F till D kontra C till A, för kravgränsen för A används elever med de preliminära betygen F till B kontra A.

Kravgränsen i kategori k erhålls genom att varje uppgift u_k inom kategorin analyseras utifrån hur den valda elevgruppen j presterat på uppgifterna. Utifrån detta beräknas sannolikheten, Angoffskattningen A_{uj} , för att en elev ur denna elevkategori får poäng på uppgiften. Kravgränserna inom varje kriteriekategori k fås, i linje med Angoffs metod, genom att Angoffskattningarna för varje uppgift inom kategorin summeras. När detta görs på hela provet erhålls den kompensatoriska kravgränsen. Matematiskt beskrivs beräkningen av kravgränserna som

$$A_{uj} = \frac{\sum_{i \in j} x_{iu}}{n_j}, \forall j \in J = \{\text{elevgrupp}\} = \{E, C, A\}$$

och summeras enligt

$$G_{kj} = \sum_{u \in k} A_{uj}, \quad \forall j \in J.$$

Där $x_{iu} \in \{1,0\}$ poäng för elev i på uppgift u , $i \in I = \{\text{elever}\} = \{E1, E2, \dots, E_n\}$, n_j är antalet elever i elevgrupp j . Vidare definieras $k \in K = \{\text{kriteriekategori}\} = \{BP, PM, KR\}$ och $u \in U = \{\text{uppgifter}\} = \{U_1, U_2, \dots, U_{25,3}\}$. Den kompensatoriska gränsen får genom formeln $G_j = \sum_u A_{uj}$, $\forall j \in J$.

Avrundningar har valts att göra uppåt endast om den uträknade kravgränsen är större än 0.75 enheter från närmaste lägre heltal, således avrundas 6.74 till 6 medan 6.75 avrundas till 7s.

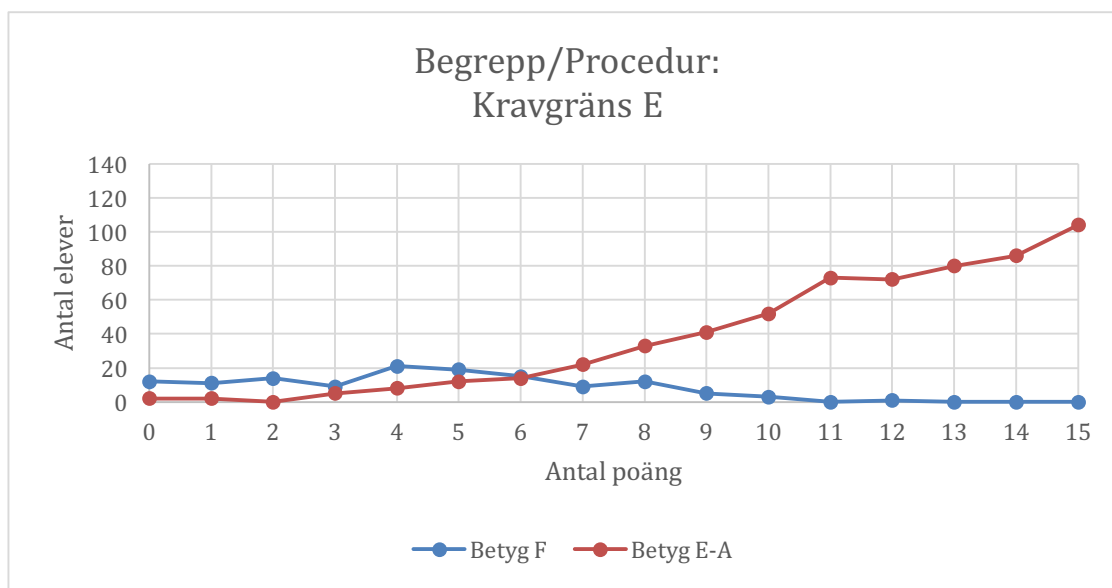
7.2.2 Vår modifierade Contrasting Groups-metod

Kravgränssättningen enligt Contrasting-Groups metoden, kräver även en expertgrupp/panel med god kunskap om provtagarnas kunskapsnivåer. För att simulera detta har metoden modifierats så att det preliminära kursbetyget som eleverna hade anger deras kunskapsnivå. På så sätt erhålls kontrasterade grupper som ligger på varsin nivå av den eftersökta kravgränsen.

För att finna kravgränsen E har provtagarna således delats upp efter vad de bedömts ha för preliminärt kursbetyg. Den ena gruppen omfattar alla provtagare med preliminärt kursbetyg F , och den andra gruppen omfattar alla provtagare med preliminärt kursbetyg E eller högre.

Därefter har de olika gruppernas prestation analyserat inom varje kriteriekategori, samt kompensatoriskt över hela provet. Kravgränsen beräknas enligt metoden som den poäng där antalet från varje grupp är lika många. Ett linjediagram – med axlarna ”Antal poäng” och ”Antal elever” – ger två linjer som representerar ”F-gruppen” respektive ”E och högre-gruppen”. Där linjerna korsar varandra är den punkten där antalet ur varje grupp är lika stora och i så fall en möjlig kravgräns. Figur 2 visar hur en sådan korsning ser ut, kravgränsen i detta fall blir således 6 poäng.

Figur 2 - Diagram som visar kravgränsen för betyg E inom kriteriekategorin Begrepp/Procedur för VT-14.



På motsvarande sätt har kravgränserna för C och A beräknats. Provtagargrupperna för kravgräns C blev, för den ena gruppen, de provtagare vars preliminära kursbetyg var D, E eller F, och för den andra gruppen, de provtagare vars preliminära kursbetyg var C, B eller A. För kravgräns A blev således den ena gruppen, de elever vars kursbetyg var B eller lägre, och för den andra gruppen, de elever vars kursbetyg var A.

I de allra flesta fall sker korsningen inte vid ett heltal, därför används samma avrundningsprincip som för Angoffs metod.

8 Resultat

Tabell 1 och Tabell 2 beskriver de kravgränser som erhålls då de valda metoderna – Contrasting Groups och Angoff – tillämpas på elevdata från de två nationella proven i Matematik 2c från VT-14 och -15. Alla diagram erhållna från Contrasting Groups metoden redovisas i Bilaga 2 och 3.

Tabell 1 De uträknade kravgränserna per kriteriekategori på provet VT14 enligt respektive metod.

Kravgränsmodell		CONTRASTING GROUPS			ANGOFF			Totalt antal poäng
		E	C	A	E	C	A	
Icke-kompensatorisk	BEGREPP/PROCEDUR	6	14	20	8	15	20	23
	PROBLEMLÖSNING/MODELLERING	2	11	18	5	12	19	24
	KOMMUNIKATION/RESONEMANG	0	2	6	1	2	5	10
Kompensatorisk		12	28	45	14	30	45	57

Tabell 2 De uträknade kravgränserna per kriteriekategori på provet VT15 enligt respektive metod.

Kravgränsmodell		CONTRASTING GROUPS			ANGOFF			Totalt antal poäng
		E	C	A	E	C	A	
Icke-kompensatorisk	BEGREPP/PROCEDUR	7	14	21	9	15	20	24
	PROBLEMLÖSNING/MODELLERING	2	8	15	3	8	15	21
	KOMMUNIKATION/RESONEMANG	1	4	8	1	4	8	12
Kompensatorisk		11	27	43	14	28	44	57

I tabellerna ovan redovisas de poängantal som utgör kravgränser inom de icke-kompensatoriska kriteriekategorierna som valts. Även en kompensatorisk kravgräns redovisas här i sista raden. Sista kolumnen beskriver totala antalet poäng inom kriteriekategorin respektive på hela provet. Ur tabellerna framgår att totalpoängen i kriteriekategorin *Kommunikation/Resonemang* är lågt, vilket även resulterar i att kravgränsen inom denna kategorin är väldigt låg. Det är vidare även noterbart att den totala poängen inom kriteriekategorin *Problemlösning/Modellering* är i samma storleksordning som den för kategorin *Begrepp/Procedur* men att kravgränserna för den tidigare kriteriekategorin är lägre än den senare.

Då de beräknade kravgränserna tillämpas på elevresultat från de två nationella prov som ingår i undersökningen kommer provbetygen fördela sig enligt Tabell 3 och Tabell 4. Tabellerna visar på en varierande skillnad i provbetyg mellan de kompensatoriska och icke-kompensatoriska metoderna. Störst skillnad erhålls av Angoffs metod för de lägre betygen, främst vad gäller provet VT-14. Vidare ses att det för betyg A är liten skillnad inom både den kompensatoriska och den icke-kompensatoriska metoden för båda modellerna. Resultaten visar även att Angoffs metod ger fler andelar elever de högsta och det lägsta betyget i förhållande till Contrasting Groups, medan den senare innehar flest elever på mittenbetygen, E och C.

Tabell 3 - Andel elever med ett visst betyg, VT-14

<i>Kravgränssättningsmetod</i>		<i>F (%)</i>	<i>E (%)</i>	<i>C (%)</i>	<i>A (%)</i>
<i>Contrasting Groups</i>	Icke-kompensatoriskt	8,5	45,8	39,0	6,7
	Kompensatoriskt	10,0	35,9	43,6	10,4
<i>Angoff</i>	Icke-kompensatoriskt	23,3	26,3	33,4	7,0
	Kompensatoriskt	13,1	38,5	38,0	10,4

Tabell 4 - Andel elever med ett visst betyg, VT-15

<i>Kravgränssättningsmetod</i>		<i>F (%)</i>	<i>E (%)</i>	<i>C (%)</i>	<i>A (%)</i>
<i>Contrasting Groups</i>	Icke-kompensatoriskt	14,0	44,8	34,6	6,5
	Kompensatoriskt	10,0	39,5	41,8	8,7
<i>Angoff</i>	Icke-kompensatoriskt	20,6	40,7	31,1	7,6
	Kompensatoriskt	14,0	38,3	37,2	10,5

9 Diskussion

Då de nationella proven, enligt förslag från Statens offentliga utredning (SOU, 2016:25), fortsättningsvis enbart kommer ha till syfte att stödja lärares betygsättning väcktes tanken om hur ett betygssystem skulle kunna sättas på nationella prov i matematik. Som det ser ut idag följer de nationella proven i matematik en kompensatorisk modell för betygsättning, där en elev kan kompensera för brister inom en förmåga med goda kunskaper inom en annan. När kursbetyg bestäms skall däremot eleven klara alla kunskapskrav på en viss nivå för att kunna erhålla detta betyg (Skolverket, 2011), det vill säga att det följer en icke-kompensatorisk modell för betygssättningen.

Syftet med denna uppsats blev sålunda att testa huruvida en icke-kompensatorisk kravgränssättningsmetod skulle kunna fungera med dagens utformning av de nationella proven i matematik. Är en sådan kravgränssättningsmetod tillämpbar på dagen nationella prov och hur skulle utfallet bli om detta gjordes?

Våra resultat visar att det är möjligt att tillämpa en sådan modell och de visar även att summan av de tre icke-kompensatoriska kravgränserna skulle bli lägre än med en kompensatorisk modell. Samtidigt skulle betygssnittet sjunka, vilket indikerar att kravgränserna, trots lägre totalpoäng, är hårdare.

9.1 Resultatdiskussion.

Utifrån resultaten är framförallt angeläget att diskutera tre områden som har med de nationella proven att göra. Det första är att poängantalet i kategorin *problemlösning/modellering* är i samma storleksordning som *begrepp/procedur*, samtidigt som den tidigare kriteriekategorin har märkbart lägre kravgränser än den senare. För det andra är de låga kravgränserna i kategorin *kommunikation/resonemang* anmärkningsvärda, och visar på svårigheter att tillämpa en icke-kompensatorisk modell på de nationella prov som undersökts. För det tredje diskuteras orsaken till att poängsumman av de icke-kompensatoriska kravgränserna inte alltid är lika med kravgränserna för den kompensatoriska.

9.1.1 Svårighetsgraden i poängfördelningen

Som tidigare nämnts är det totala antal poäng i kategorin *problemlösning/modellering* och *begrepp/procedur* av samma storleksordning. Kravgränserna följer dock inte samma trend. För VT-15 ses en markant skillnad mellan kravgränserna för dessa kriteriekategorier. För VT-14 är skillnaden inte lika tydlig, men kravgränserna för *problemlösning/modellering* är ändå lägre än den för *begrepp/procedur*, trots att totala antalet poäng är högre för den tidigare kategorin. Detta kan ha sin grund i det faktum att poängantalet inom svårighetsnivåerna i kategorin *Begrepp/Procedur* är fallande, medan för *Problemlösning/Modellering* är poängantalet istället jämnt fördelat eller ökande (se Bilaga 1). Den ojämna fördelningen av nivåpoäng kan även ses i kriteriekategorin *kommunikation/resonemang*.

En konsekvens av detta är att då mestadels höga nivåpoäng för kriteriekategorin *problemlösning/modellering* finns tillgängliga, kommer kravgränsen för framförallt betyg E bli låg. En liten bank möjliga uppgifter som en elev med precis godkänd kunskapsnivå rimligtvis kan ta poäng på, kan resultera i att denna kategori uppfattas särskilt svår och således blir en slags ”spärr” för att erhålla ett visst betyg.

Vidare kan även reliabiliteten utifrån en precis godkänd elevs synvinkel diskuteras. Antalet uppgifter en sådan elev har möjlighet att plocka poäng på kan anses låg, vilket gör att utifrån dennes perspektiv finns enbart dessa uppgifter. Få uppgifter medför en lägre reliabilitet. (Heartel, 2006, s. 75)

9.1.2 Totalt antal poäng

Det är också värt att reflektera till de låga kravgränser för *kommunikation/resonemang* som är resultatet av vår studie. Detta gäller för både VT14 respektive VT15 och de båda kravgränssättningsmetoderna. Speciellt gäller det kravgränsen för E, som vid tre utav fallen hade **en** poäng som kravgräns och i det sista fallet inte krävde någon poäng alls. De låga kravgränserna har sitt ursprung dels av samma anledning som diskuterats ovan – det vill säga att svårighetsgraden av poängen inom denna kategori är högre - dels i det låga antalet möjliga poäng inom kategorin. Då en kravgräns sätts till noll poäng kan det även ses som att denna kriteriekategori inte testas, vilket enbart sänker reliabiliteten i förhållande till sin kompensatoriska motsvarighet.

Trots den låga kravgränsen för kategorin *kommunikation/resonemang*, kan i Tabell 3

och Tabell 4 ses att den icke-kompensatoriska kravgräns-modellen är hårdare i sina bedömningar. Andelen elever som fått de högre betygen är nämligen större för den kompensatoriska modellen jämfört med den icke-kompensatoriska, vilket även är förväntat (Haladyna & Hess, 1999). Dock ses för provet VT-14 inom Contrasting Groups-metoden att den kompensatoriska modellen underkänner en större andel än den icke-kompensatoriska. I beaktande måste dock tas att det totala antalet poäng för betyg E med den kompensatoriska Contrasting Groups-metoden är avsevärt högre än dess icke-kompensatoriska motsvarighet, hela 50% större.

9.1.3 Kompensatoriska kontra icke-kompensatoriska kravgränser

När en jämförelse görs mellan kravgränserna för de olika modellerna ses att de totala poäng som krävs inom de icke-kompensatoriska metoderna, inte skiljer sig markant från respektive kompensatoriska motparter, förutom för kravgräns E för Contrasting Groups VT-14. I de flesta fall har de kompensatoriska metoderna högre kravgränser än de totala antal poäng som de icke-kompensatoriska motparterna kräver. Trots detta faktum kan i Tabell 3 och Tabell 4 ses att under en icke-kompensatorisk modell är betygsättningen hårdare, vilket överensstämmer med tidigare forskning (Haladyna & Hess, 1999).

En fråga som väcks är varför de kompensatoriska kravgränserna **inte** är lika med summan av kravgränserna per kriteriekategori i respektive icke-kompensatorisk modell. I de allra flesta fall

kan detta förklaras med de avrundningar som gjorts då korsningarna mellan graferna bestämts. Dock finns en problematisk kravgräns, den för *Contrasting Group VT-14*, där det skiljer fyra poäng mellan den kompensatoriskt satta kravgränsen och summan av kravgränserna på den icke-kompensatoriska. Den uppsatta policyn för hur avrundningar sker slår hårdare mot de icke-kompensatoriska metoderna då det där sker tre avrundningar istället för en. En del i dessa fyra poäng går att förklara med avrundning, men inte fullt ut. Dock har vi inte hittat någon anledning till att differensen blir så stor.

9.2 Metoddiskussion

För att besvara forskningsfrågorna valdes två kravgränssättningsmetoder. Dessa val förklaras närmare i 7.2 *Val av kravgränssättningsmetoder*. *Contrasting Groups*- metoden utgår från lärares preliminära betyg på eleverna vilket resulterar i ett interbedömmarreliabilitetsproblem – vilket innebär att samstämmigheten mellan lärarnas bedömningar bör vara hög (Wikström, 2013) – samt en viss grad av osäkerhet gällande lärares engagemang kring detta preliminära betyg. Det går även att argumentera för att det som faktiskt mäts när metoden appliceras är just lärares förmåga att sätta dessa preliminära kursbetyg, som kan tänkas att det görs på ett varierande betygsunderlag.

När den modifierade versionen av Angoffs metod appliceras utgår det från en expertgrupps utlåtande vad gäller kravgränserna. Här skall tas i beaktande att de kravgränser som finns för betygen C och A utöver ett visst poängantal även kräver poäng på en viss nivå. Detta tas inte i beaktande när kravgränserna för våra modeller beräknas, utan enbart den totala poängen används. Dock betonar vi att det gjordes både för den icke-kompensatoriska som för den kompensatoriska varianten av Angoffs metod.

Problematiken som kan uppfattas vid användandet av kravgränssättningsmetoderna försvagas dock då vi använder två metoder, som är tämligen olika. Det vi undersöker är inte effektiviteten av kravgränssättningsmetoder utan olika kravgränsmodeller med hjälp av kravgränsmetoder och utifrån det fås likartade resultat. Detta styrker snarare resonemangen kring applicerandet av kravgränserna.

Som ytterligare metod hade en holistisk varit intressant att studera, då undersökningen redan omfattar en analytisk och en provtagarcentrerad metod. Detta har av tidsskäl inte varit möjligt, liksom att datamängden hade begränsningar. Många, i det närmaste alla, metoder använder sig även av en expertgrupp vilket vi inte hade tillgång till utan har förlitat oss på den information kring olika kompetenta gruppers omdöme, som finns tillgänglig. Att tillsätta en expertgrupp för att göra en holistisk bedömning hade inte bara ur tidsaspekten varit omöjlig, utan hade även sänkt interbedömmarreliabiliteten mellan de tre olika expertgrupperna.

En ytterligare fråga som väcks är om vi hade kunnat styrka våra resonemang genom att applicera de två valda kravgränssättningsmetoderna på ytterligare prov. Det hade varit möjligt, men av tidsskäl gjordes inte detta. Det skall även tilläggas att enbart sju prov i matematik 2c finns tillgängliga som är anpassade för gy-11, vilket är den läroplan som är utgångspunkt för

undersökningen. De två senaste prover som gjort är de som används i undersökningen, detta för att provkonstruktionen skall hinna bli stabilare och således mer jämförbar.

De nationella proven är idag inte utformade för att vara icke-kompensatoriska varför det visar sig problematiskt att applicera ett sådant kravgränssystem på provet. Bland annat sänks reliabiliteten snabbt sänks då provet delas in i delar. Storleken på dessa delar kan ses som en andel av poängen på hela provet, vilket gör att kriteriekategorin *begrepp/procedur* innehåller ca 60% av hela antalet poäng, *problemlösning/modellering* innefattar ca 42% och de återstående ca 17% återfinns i kriteriekategorin *kommunikation/resonemang*. Reliabiliteten på dessa nya kriteriekategorier beror på reliabiliteten på dagens nationella prov, vilket inte är uppgifter vi fått tillgång till. Det som dock kan sägas är att reliabiliteten per kriteriekategori kommer att vara i storleksordningen av drygt hälften av andelen poäng i delkategorin. Exempelvis i *begrepp/procedur* kan reliabiliteten, enligt *Spearman-Browns prophecy formula* (Heartel, 2006), **maximalt** vara 43% medan reliabiliteten för *problemlösning/modellering* maximalt är 27%. Dessa siffror är beräknade utifall reliabiliteten på dagens nationella prov skulle vara 100% vilket inte är fallet. Siffrorna kan dock ses som riktmärken på storleksordningen av reliabilitetssänkningen.

9.3 Slutsatser och vidare forskning

Vi drar utifrån ovanstående diskussioner slutsatsen att det nationella provet idag inte är anpassat för att användas i en icke-kompensatorisk modell vid betygsättning. För att en sådan modell skall fungera krävs fler poäng så att alla kriteriekategorierna får kravgränser som säger något om vilken nivå som eleven presterar på. Samtidigt måste även svårighetsgraden inom varje kriteriekategori vara jämnare fördelad. Vidare krävs även att fortsatt hög reliabilitet bibehålls på provet, något vi inte undersökt. Vi lämnar till vidare forskning att studera hur ett sådant prov skulle kunna utformas, om det ens är möjligt, samt vilken kravgränssättningsmetod som bäst lämpar sig för detta prov.

Bakgrunden till vår undersökning ligger i att det finns en motsättning i hur kursbetyg och provbetyg sätts. Då denna motsättning kvarstår bör antingen betygssystemet ändras, till att återigen vara kompensatoriskt, eller proven omarbetas till att vara icke-kompensatoriska.

Med bakgrund i vår undersökning föreslår vi att, om proven omarbetas, skall betyg inte sätts grundade på skarpa kravgränser utan inom ett intervall, exempelvis 10-12 poäng. Nästa förslag är att det kompensatoriska betyget tas bort och att tre eller fler betyg bestäms, exempelvis utifrån de kriteriekategorier som används i undersökningen. Vidare anser vi att kriteriekategorin *Kommunikation/Resonemang* inte är lämplig att bedömas med enskilda poäng satta på uppgifter, utan föreslår istället en holistisk bedömning över hela provet.

10 Referenslista

- Cizek, G. J., & Bunch, M. B. (2007). *Standard Setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA, USA: Sage Publications.
- Crocker, J., & Algina, L. (2008). *Introduction to classical & modern test theory*. Mason, Ohio, USA: Cengage Learning.
- Gustafsson, J.-E., Cliffordson, C., & Erickson, G. (2014). *Likvärdig kunskapsbedömning i och av den svenska skolan - problem och möjligheter*. Stockholm: SNS.
- Haladyna, T., & Hess, R. (1999). An Evaluation of Conjunctive and Compensatory Standard-Setting Strategies for Test Decisions. *Educational Assessment*, 6(2), ss. 129-153. doi:10.1207/S15326977EA0602_03
- Hambleton, R. K. (1995). Setting Standards on Performance Assessments: Promising New Methods and Technical Issues. *The Annual Meeting of the American Psychological Association* (s. 12). New York, USA: THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC). Hämtat från <http://eric.ed.gov/?id=ED403289>
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. i R. L. Brennan, *Educational Measurement Fourth Edition* (ss. 433-471). Westport, CT, USA: American Council on Education and Preager Publishers.
- Heartel, E. H. (2006). Reliability. i R. L. Brennan, *Educational Measurement Fourth edition* (ss. 65-110). Westport, CT, USA: American Council on Education and Preager Publishers.
- Livingston, S. A., & Zieky, M. J. (1982). *Passing Scores: A manual for setting standards of performance on educational and occupational tests*. Educational Testing Service.
- Mehrens, W. A., & Cizek, G. J. (2012). Standard Setting for Decision Making: Classifications, Consequences, and the Common Good. i G. J. Cizek (Red.), *Setting Performance Standards - Foundations, Methods, and Innovations* (2:a uppl., ss. 33-46). New York, USA: Routledge.
- Skolverket. (2011). *Läroplan, examensmål och gymnasiegemensamma ämnen för gymnasieskola 2011*. Hämtad från: <http://www.skolverket.se/publikationer?id=2705>.
- Skolverket. (2012). *Bedömning och betygsättning i gymnasieskolan*. <http://www.skolverket.se/publikationer?id=2841>.
- Skolverket. (2016a). *Kommentarer till ämnesplan. Ämne - Matematik*. Hämtat från 2016-05-16 http://www.skolverket.se/laroplaner-amnen-och-kurser/gymnasieutbildning/gymnasieskola/mat?tos=gy&subjectCode=MAT&lang=sv&courseCode=MATMAT02c#anchor_MATMAT02c.
- Skolverket. (2016b). *Relationen mellan provbetyg och betyg*. Hämtat från <http://www.skolverket.se/bedomning/betyg/provbetyg>
- Skolverket. (2016c). *Tidigare givna prov*. Hämtat från <http://www5.edusci.umu.se/np/resources/libraries/track.php?file=/np/np-2-4-prov/Ma2c-vt15.pdf>
- SOU. (2016:25). *Likvärdigt, rättssäkert och effektivt-ett nytt nationellt system för kunskapsbedömning*. Stockholm: Elanders Sverige AB.
- Umeå Universitet. (2016). *Nationella kursprov i matematik 2-4*. Hämtat från Institutionen för tillämpad utbildningsvetenskap: <http://www.edusci.umu.se/np/np-2-4/tidigare-givna-prov/>
- Wikström, C. (2013). *Konsten att göra bra prov - vad lärare behöver veta om kunskapsmätning*. Stockholm: Natur och kultur.

Bilaga 1 – Poängfördelning VT-15

Poängfördelningen på provet från VT-15 (Skolverket, 2016c). Då provet från VT-14 inte är frisläppt kan dessa poäng och poängfördelningar inte redovisas, men formatet följer samma princip.

NpMa2c vt 2015

Provsammanställning – Kunskapskrav

Tabell 1 Kategorisering av uppgifterna i kursprovet i Matematik 2c i förhållande till nivå och förmågor. Poängen i denna tabell anges i samma ordning som i bedömningsanvisningen. Till exempel motsvarar 4_1 och 4_2 den första respektive andra poängen i uppgift 4.

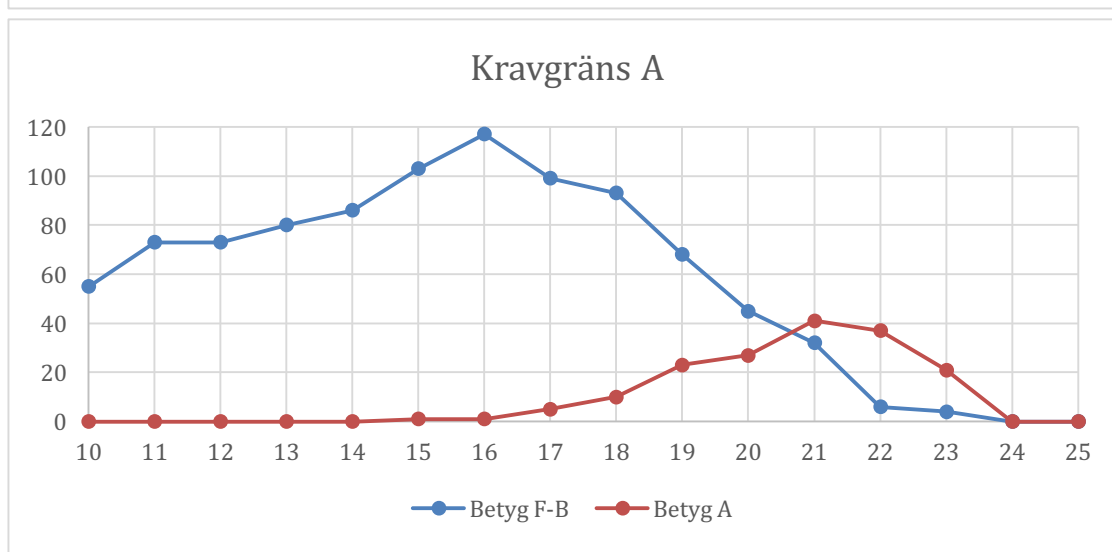
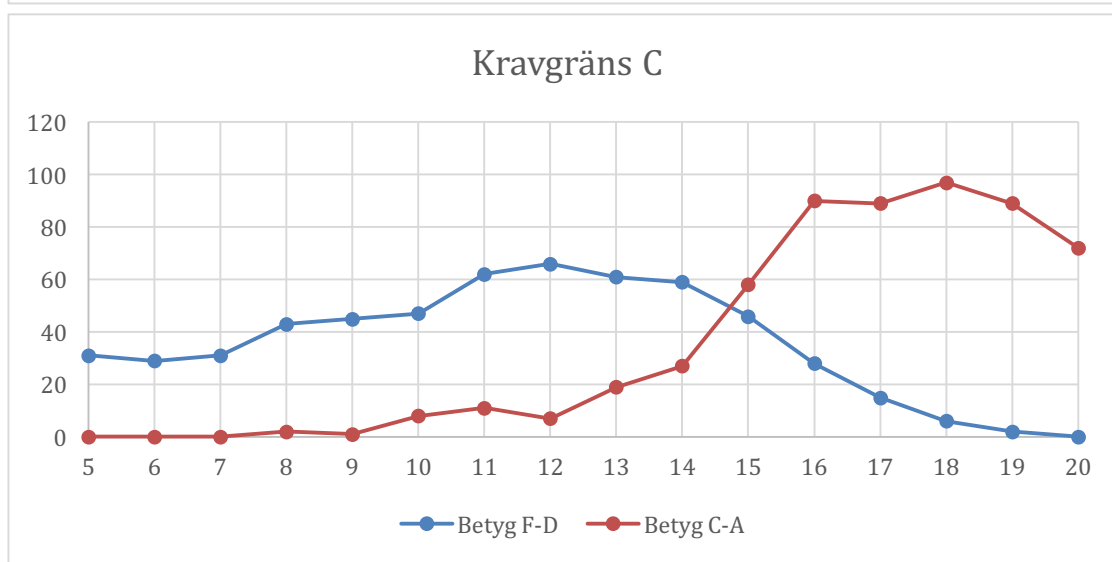
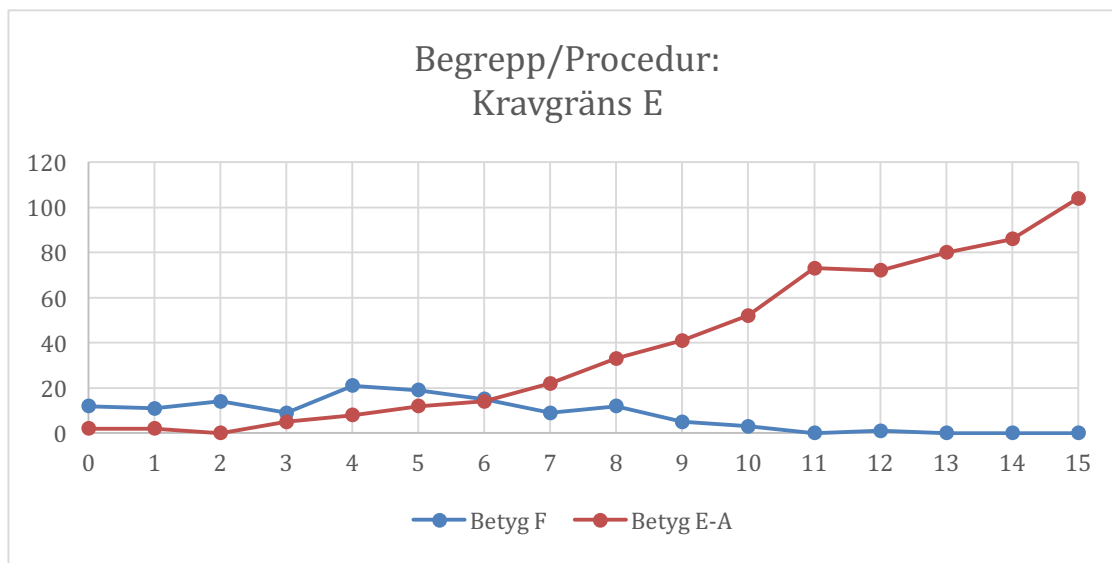
Delprov	Uppg. Poäng	Förmåga och nivå																
		E				C				A								
		B	P	PM	RK	B	P	PM	RK	B	P	PM	RK					
B	1		1															
	2a		1															
	2b		1															
	3a		1															
	3b			1														
	4_1	1																
	4_2	1																
	5						1											
	6						1											
	7							1										
	8a_1						1											
	8a_2							1										
	8b								1									
	9a									1								
	9b										1							
	C	10_1		1														
		10_2		1														
11_1			1															
11_2			1															
12_1				1														
12_2							1											
12_3							1											
13_1							1											
13_2							1											
14_1								1										
14_2								1										
15_1									1									
15_2									1									
16_1																1		
16_2																1		
16_3																1		
17a_1								1										
17a_2								1										
17b_1															1			
17b_2															1			
17b_3															1			

Delprov	Uppg. Poäng	Förmåga och nivå																
		E				C				A								
		B	P	PM	RK	B	P	PM	RK	B	P	PM	RK					
D	18_1									1								
	18_2									1								
	19_1									1								
	19_2									1								
	20_1	1																
	20_2	1																
	21a_1										1							
	21a_2										1							
	21b_1										1							
	21b_2										1							
	22_1											1						
	22_2											1						
	22_3												1					
	23_1													1				
	23_2															1		
	24_1															1		
	24_2															1		
	25_1															1		
	25_2															1		
	25_3															1		
	25_4															1		
	Total	4	8	6	2	3	5	7	5	2	2	8	5					
	Σ	57	20				20				17							

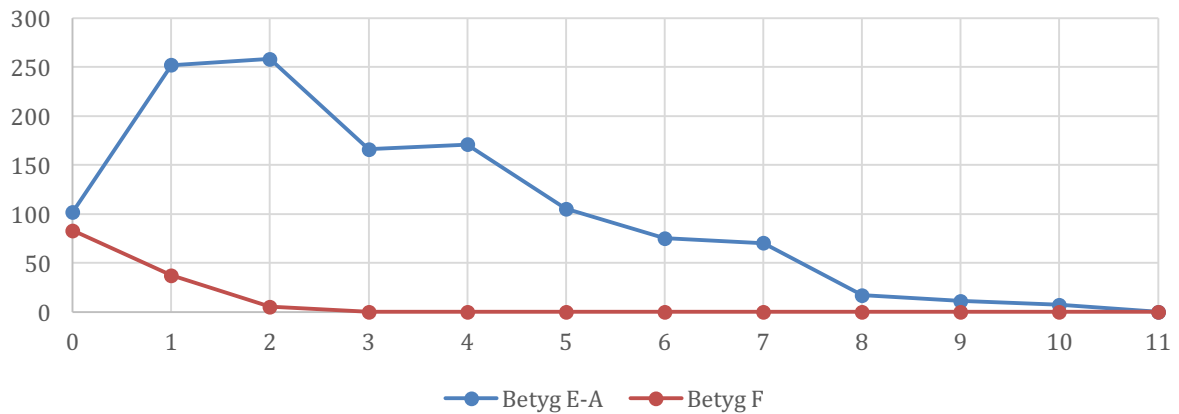
B = Begrepp, P = Procedur, PM = Problemlösning/Modellering och RK = Resonemang/Kommunikation

Bilaga 2 – Diagram för kravgränser VT-14

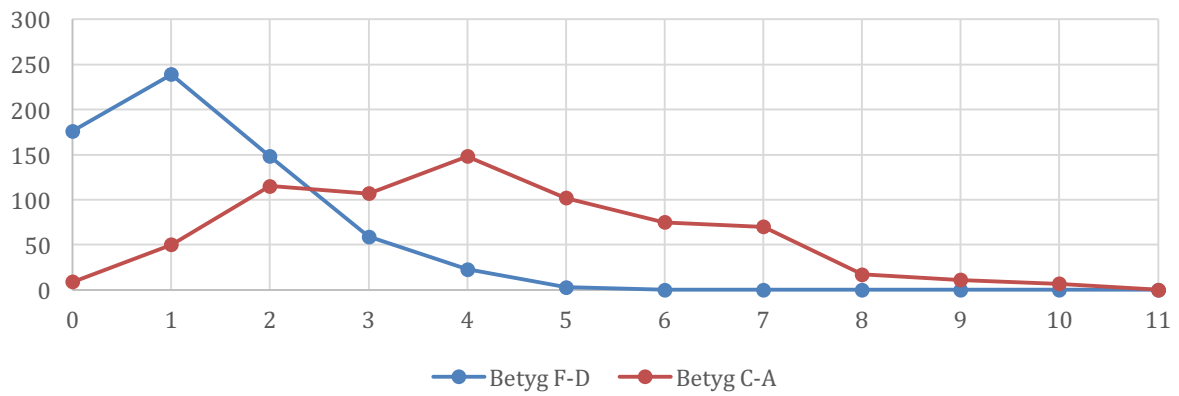
Här redovisas alla grafer som använts för att räkna ut kravgränserna på provet VT-14. x-axeln beskriver antal poäng och y-axeln beskriver antal elever med detta poäng.



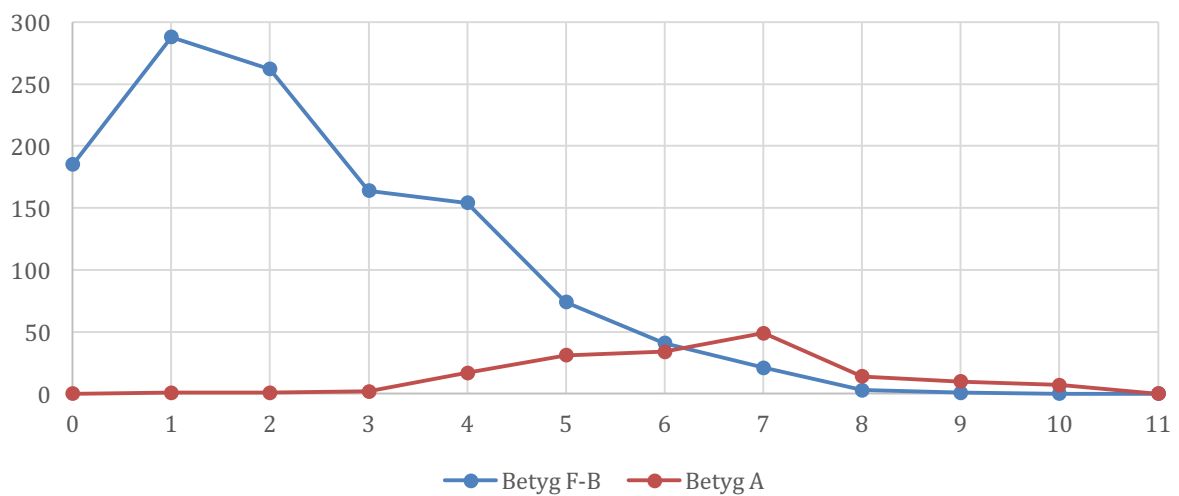
Kommunikation och Resonemang: Kravgräns E



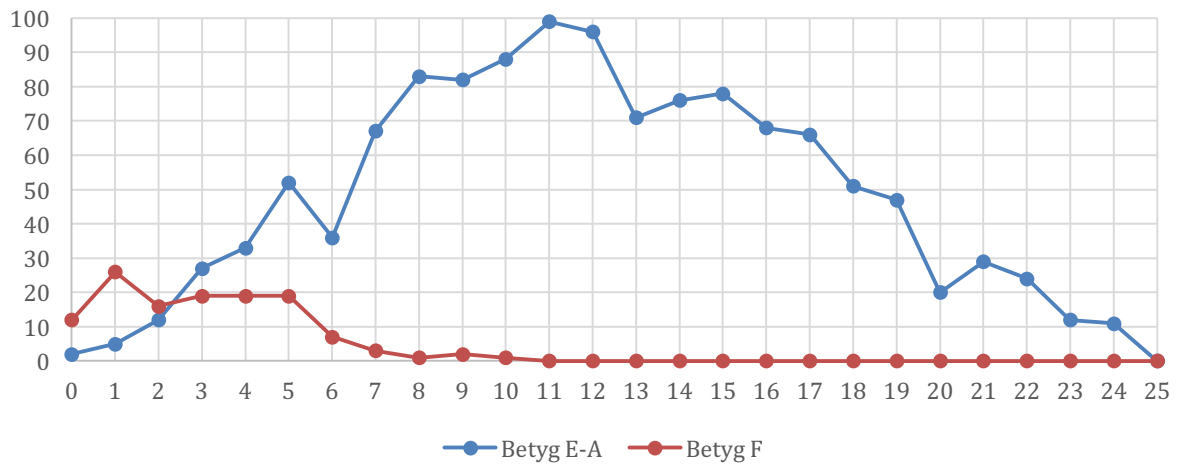
Kravgräns C



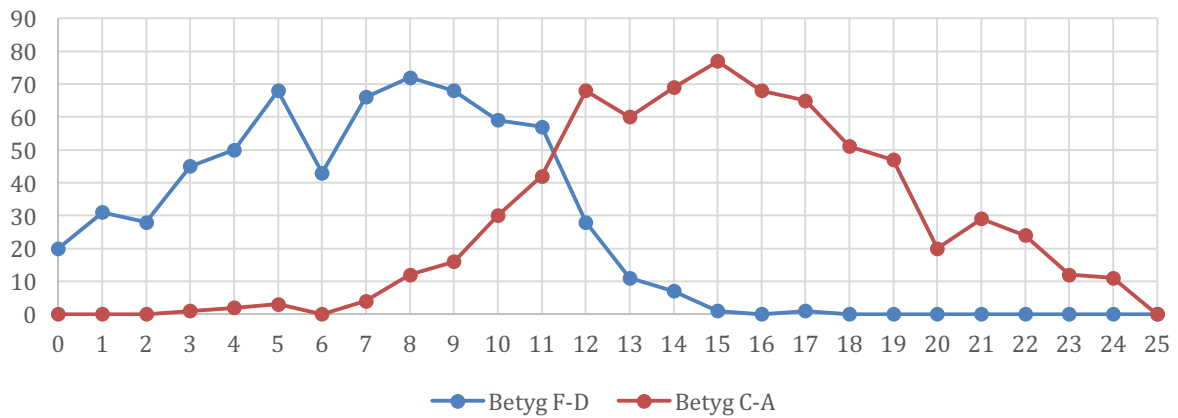
Kravgräns A



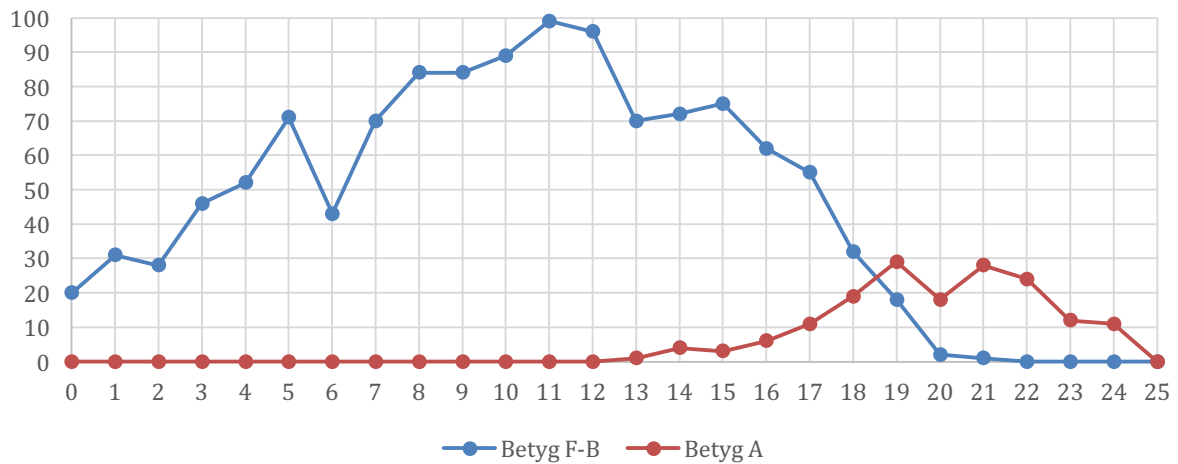
Problemlösning/Modellering: Kravgräns E



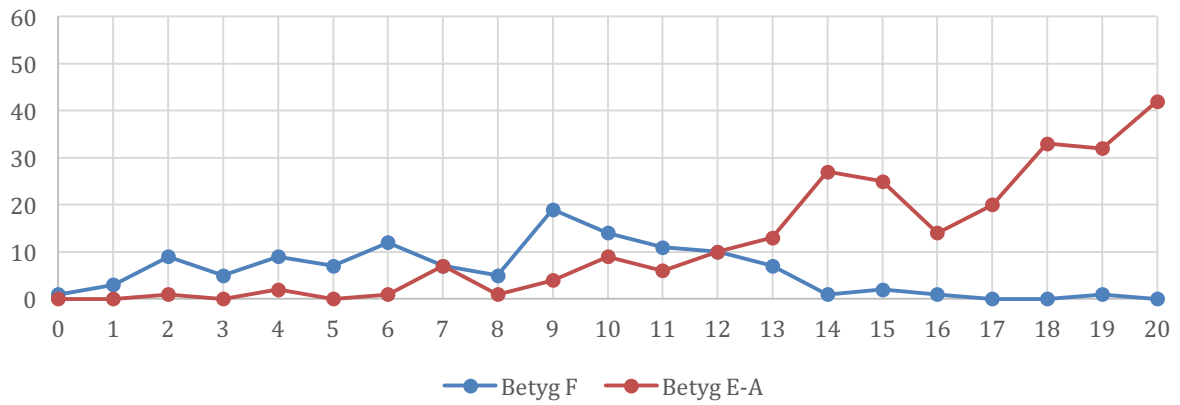
Kravgräns C



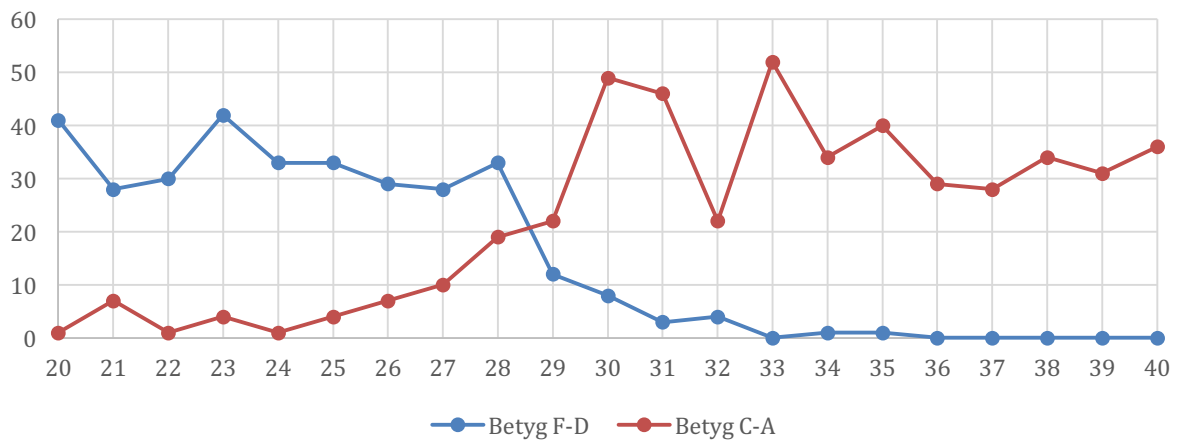
Kravgräns A



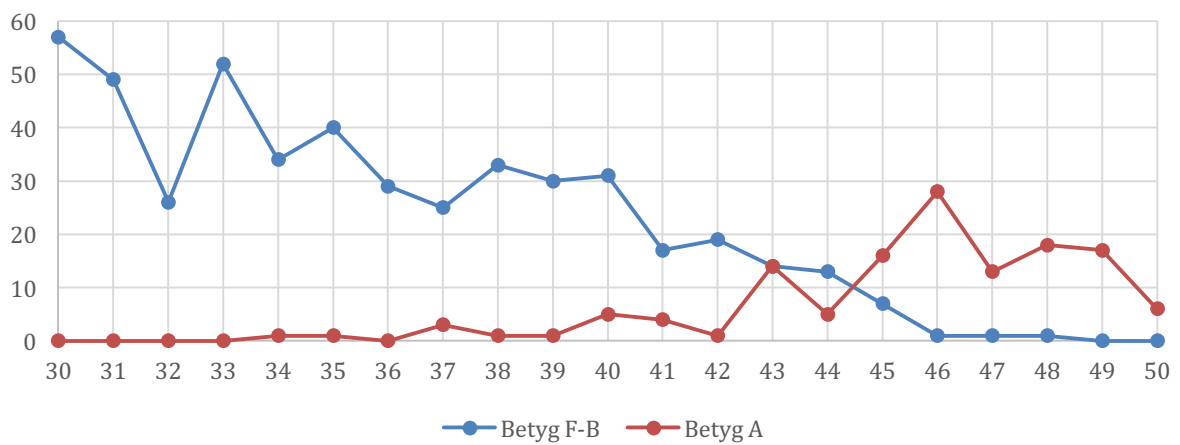
Kompensatoriskt: Kravgräns E



Kravgräns C

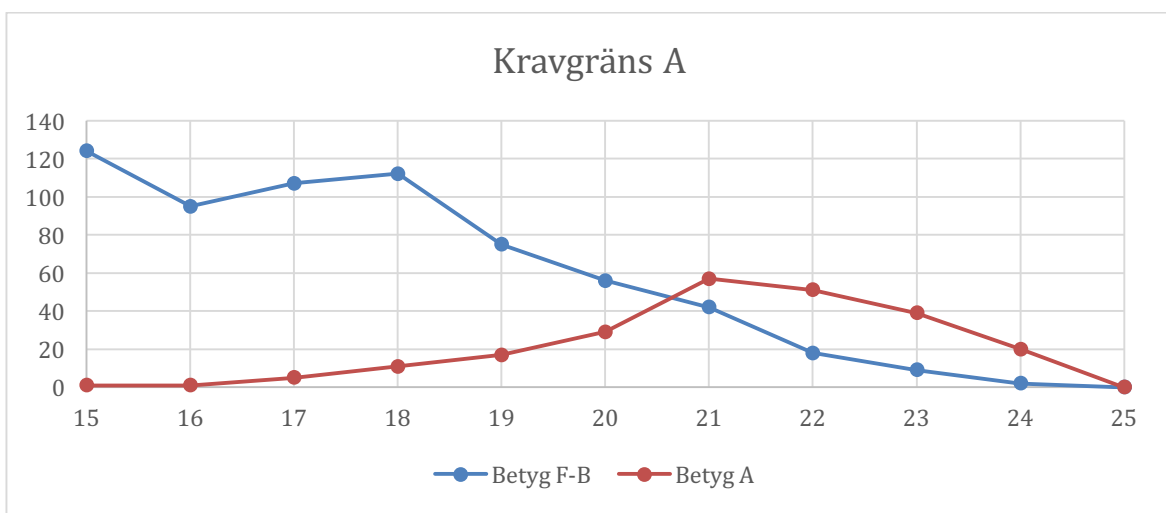
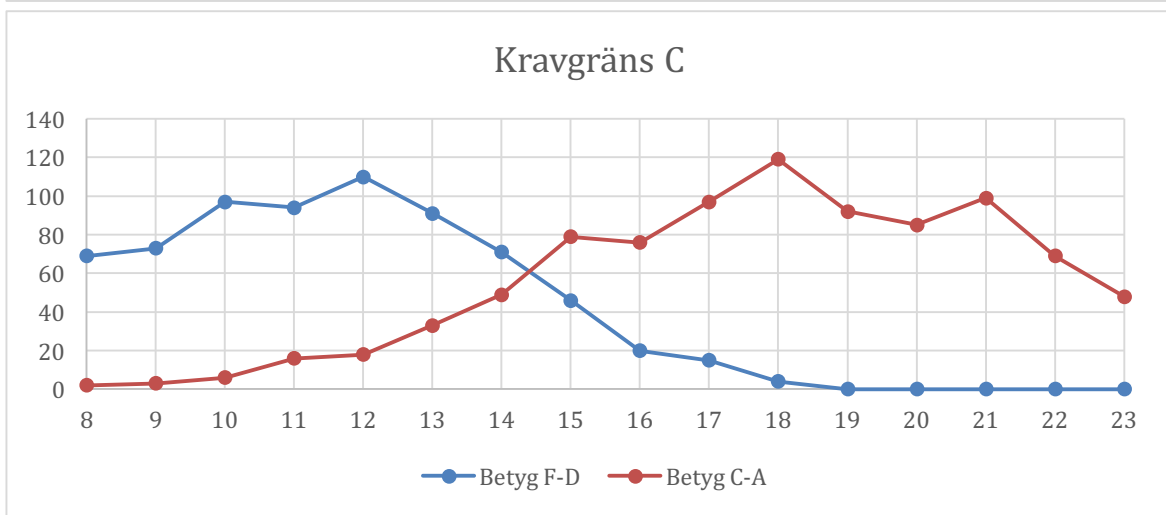
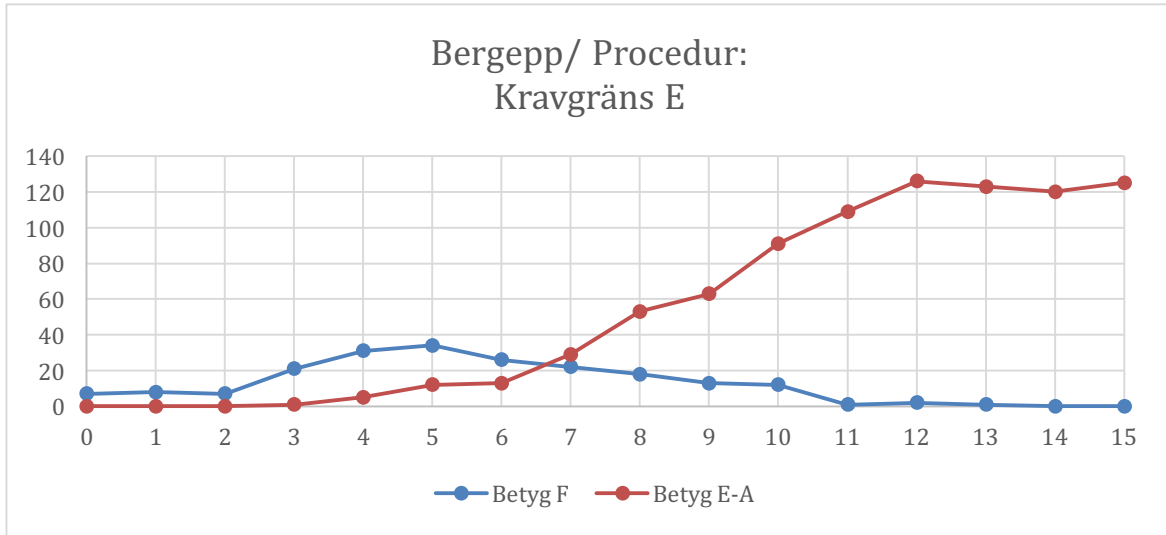


Kravgräns A

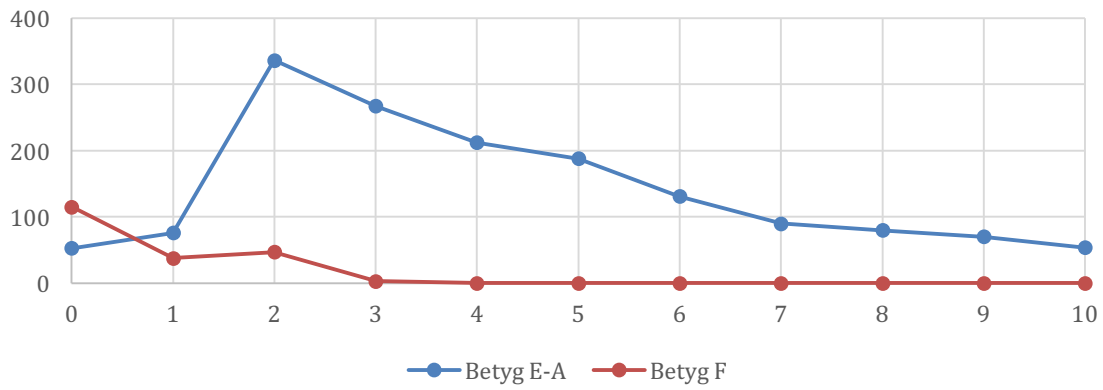


Bilaga 3 – Diagram för kravgränser VT-15

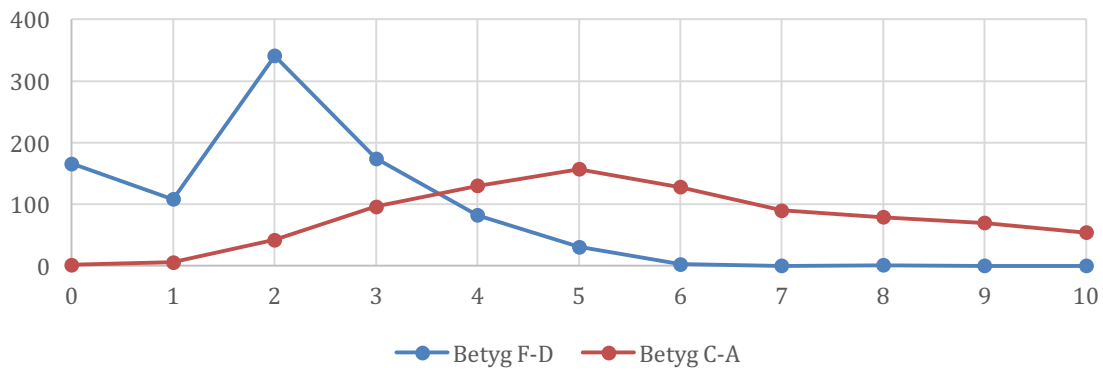
Här redovisas alla grafer som använts för att räkna ut kravgränserna på provet VT-15. Antal poäng beskrivs på x-axeln medan y-axeln beskriver antal elever med detta poäng.



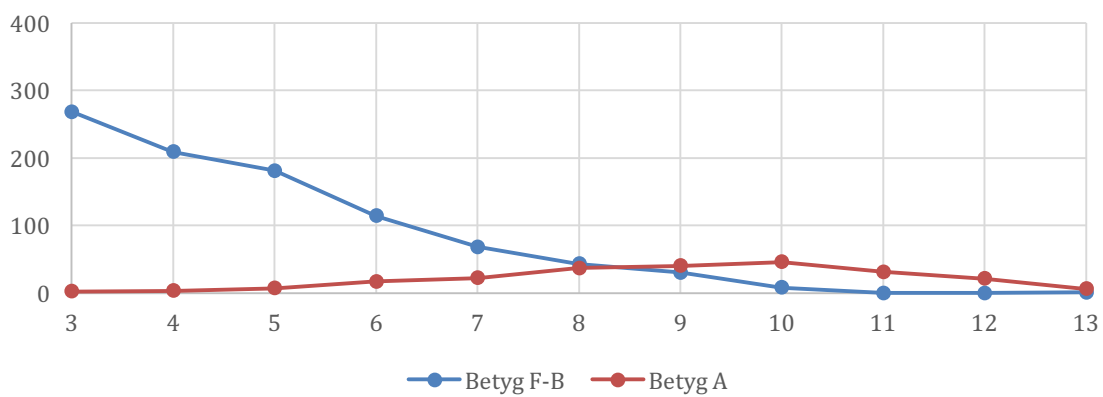
Kommunikation/Resonemang: Kravgräns E



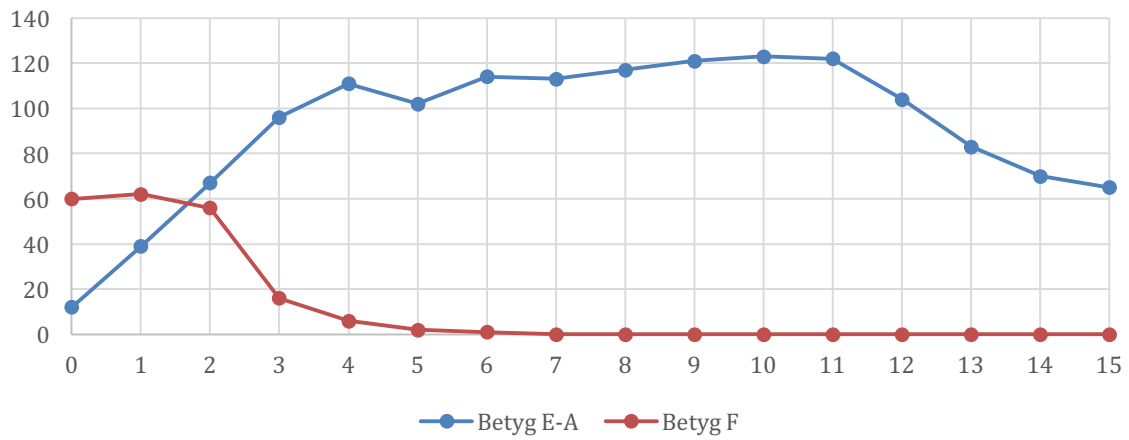
Kravgräns C



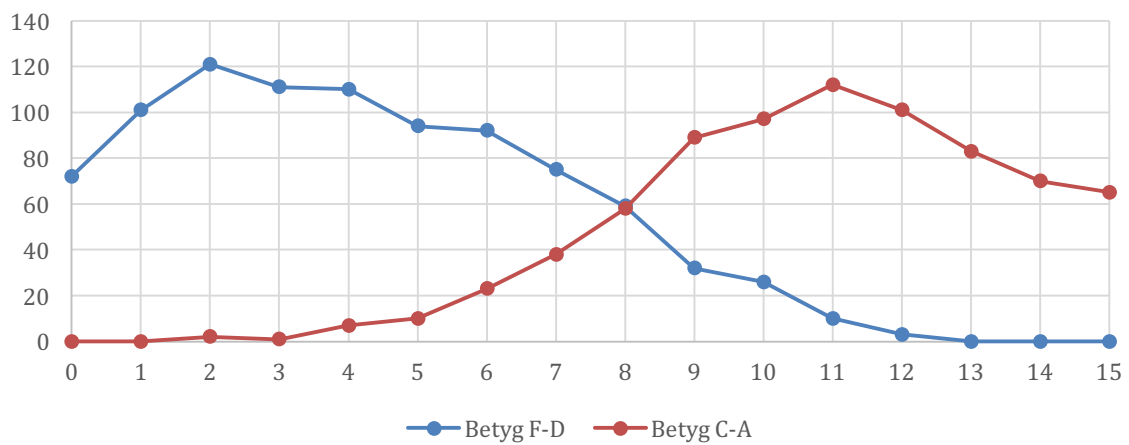
Kravgräns A



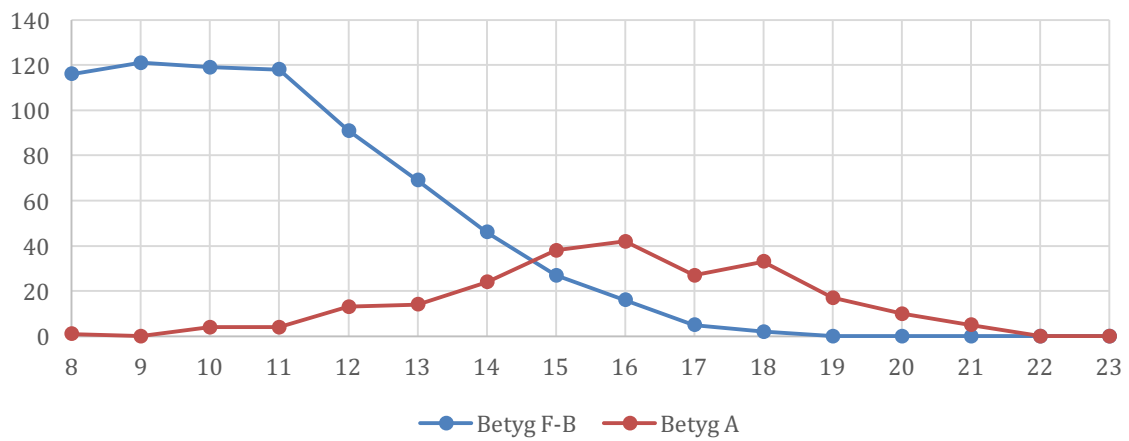
Problemlösning/Modellering: Kravgräns E



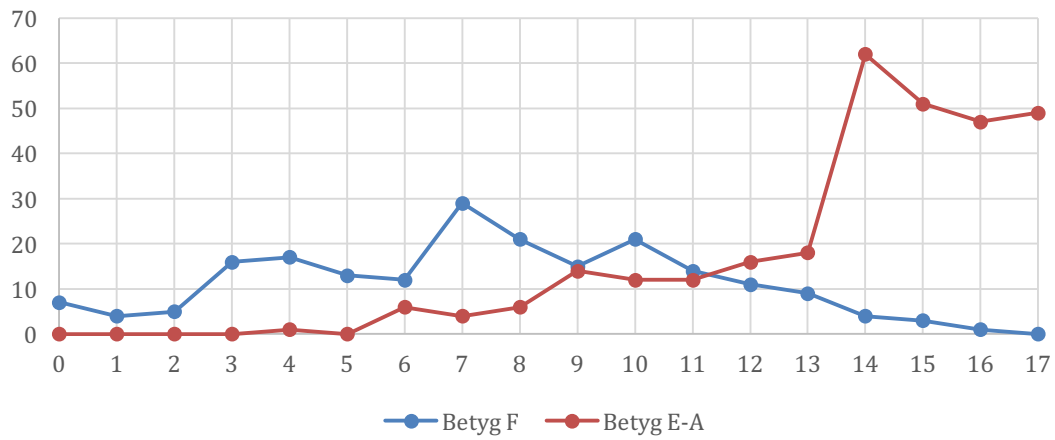
Kravgräns C



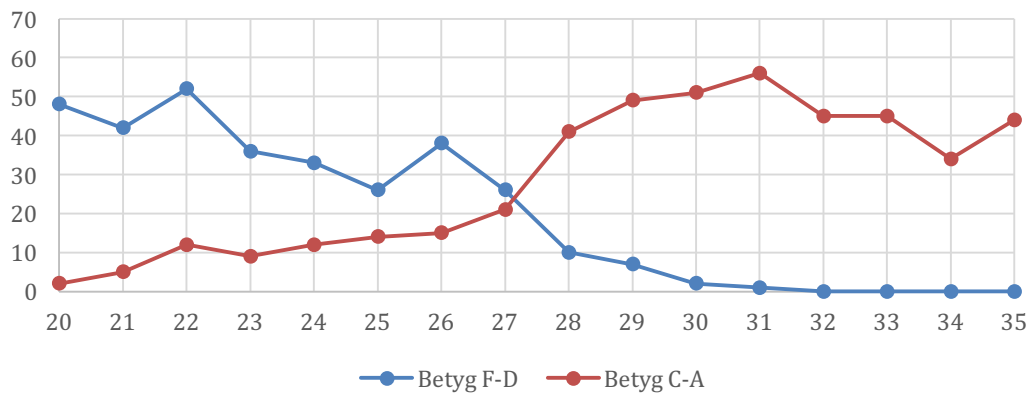
Kravgräns A



Kompensatoriskt: Kravgräns E



Kravgräns C



Kravgräns A

