

GÖTEBORGS UNIVERSITET
Institutionen för svenska språket

Rättssäker bedömning?
En fallstudie på Invandrarakademins slutrapport

Emma Almingefeldt

Självständigt arbete, 15 hp
Svenska som andraspråk, SSA136
Vårtermin 2013
Handledare: Rakel Johnson

Sammandrag

Invandrarakademien är en högskoleförberedande kurs vid Högskolan i Borås som ger nationell behörighet i Svenska 2/Svenska som andraspråk 2 (och Samhällskunskap A) efter godkänd kurs. Sista examinationen som genomförs av studenterna utgörs av skrivandet av en akademisk rapport. Uppsatsens syfte är att undersöka om Invandrarakademins slutrapport kan anses vara giltig som examinationsform i ämnet svenska som andraspråk på gymnasienivå. Validiteten analyserades utifrån tre olika bedömningsmodeller; Messicks modell, Kanes tolkningskedja, Bachman och Palmers testkvaliteter samt språkskalan GERS.

Materialet som analyserades utifrån ovannämnda modeller är de skriftliga och muntliga instruktioner som finns tillgängliga för studenterna på lärplattformen. Detta material kompletterades med två intervjuer med examinatorerna som också är testkonstruktörer, handledare och undervisande lärare.

Slutsatsen är att examinationen anses hålla en relativt hög giltighet och tillförlitlighet baserat på resultaten som framkom i valideringsstudien.

Nyckelord:

Svenska som andraspråk, bedömning, Gemensam europeisk referensram för språk, Kanes tolkningskedja, Messicks giltighetsmodell, Bachman och Palmers testkvalitéer

Innehåll

1. Inledning	1
1.1. Syfte	3
1.2. Avgränsningar	3
1.3. Disposition	3
2. Bakgrund - Invandrarakademin	3
2.1 Invandrarakademins kursplan	4
3. Teori och modeller	5
3.1 Bedömning	5
3.2 Gemensam europeisk referensram för språk (GERS)	6
3.3 Kanes tolkningskedja	9
3.4 Messicks giltighetsmodell	10
3.5 Bachman och Palmers testkvalitéer	12
4. Metod och material	13
4.1 Material och metod	13
4.2 Etiska principer	14
4.3 Giltighet och tillförlitlighet	14
5. Resultat	15
5.1 Rättssäker bedömning	15
5.2 Kanes tolkningskedja	16
5.3 Messicks giltighetsmodell	17
5.4 Backman och Palmers testkvaliteter	19
5.5 Gemensam europeisk referensram för språk (GERS)	20
6. Diskussion och slutsats	21
6.1 Resultatdiskussion	21
6.2 Metoddiskussion	24
6.3 Slutsats	25
Litteratur	
Bilaga 1 - intervjufrågor	

Figurförteckning

Figur 1: Messicks modell (efter McNamara & Roever 2006:13)

Figur 2: Invandrarakademins slutuppgift i Messicks modell

1. Inledning

Bedömning är en stor del av en lärares uppdrag och bedömningarna som görs påverkar studenterna i mindre eller större utsträckning, men kan också få inverkan på människorna runt omkring dem. Denna uppsats som du läser nu har blivit bedömd liksom alla andra uppsatser som du läst. Uppsatser som denna liksom alla examinationer som görs vid ett universitet eller högskola i Sverige är bedömda av en statligt anställd myndighetsperson som skall förhålla sig till första kapitlet och den nionde paragrafen i Regeringsformen (RF), vilken är en av Sveriges grundlagar.

Domstolar samt förvaltningsmyndigheter och andra som fullgör uppgifter inom den offentliga förvaltningen skall i sin verksamhet beakta allas likhet inför lagen samt iakttaga saklighet och opartiskhet. (Regeringsformen, kap 1, §9)

En rättssäker bedömning skall således göras objektivt och fokusera på innehåll och inte på person. Detta är dock lättare i teorin än det är i praktiken.

Jag har erfarenhet av kollegor som har godkänt invandrade läkare på C1-nivå (Gemensam europeisk referensram för språk) där andra kollegor har ansett dem vara på högst B2-nivå. Förmodligen gjordes den förstnämnda bedömningen för att godkännandet var avgörande för om dessa studenter skulle få påbörja AT-tjänst eller inte. De möjliga följderna av ett godkänt eller inte påverkar inte bara individen (mikro), utan också gruppen (makro). I fallet ovan skulle kursens trovärdighet kunna påverkas ifall studenter ”kan glida igenom” och att detta sprids till diverse instanser. I exemplet ovan är det lätt att se konsekvenserna för samhället, det vill säga att samhället får ännu tillgång till en duktig välutbildad läkare som något annat land har fått bekostat utbildningen för. Konsekvenserna kan dock bli katastrofala om det uppstår språkliga missförstånd mellan läkaren och patienten och/eller med resten av vårdteamet. Bedömning är således inte bara något som skall ses utifrån ett mikroperspektiv utan också utifrån ett makroperspektiv.

I min roll som universitetsadjunkt i svenska som andraspråk vid Högskolan i Borås bedömer jag kontinuerligt mina studenter i olika slags summativa och formativa tester tillsammans med mina kollegor. Våra studenter är invandrade akademiker och de bedöms av oss

huruvida de har nått C1-nivå i svenska som andraspråk vid slutförd kurs eller inte. *Gemensam europeisk referensram för språk* (GERS) är den nivåskala som studenterna bedöms utifrån och denna referensram är inskriven i kursplanen. Kursens namn är *Invandrarakademin, 60 hp*, och den finns med i bedömningshandboken, vilket betyder att kursen likställs med andra högskoleförberedande utbildningar. Studenter med godkända betyg kan få sina utländska utbildningar validerade och de blir behöriga för kompletterande utbildningar på högskola/universitet.

För många studenter avgörs en del av deras fortsatta karriär av detta betyg, exempelvis krävs C1 för att få lärarbehörighet (enligt Skolverket) eller AT-tjänst som läkare i Sverige (enligt Socialstyrelsen). Detta ställer höga krav på oss lärare, eftersom vi måste testa det som vi faktiskt skall testa enligt kriterierna för C1 och att detta görs rättssäkert, alltså opartiskt och sakligt.

När en uppgift bedöms av lärare i skolans värld är det flera faktorer som påverkar bedömningen både ur ett makro- och ett mikroperspektiv. Dessa faktorer är dessutom både explicita och implicita. Kursmålen för universitetsstudenter och kurskriterier för grundskolan och gymnasieskolan är explicita mål som studenterna skall bedömas utifrån, medan till exempel IT-kunskaper kan vara implicit på grund av att testet görs med hjälp av ett IT-verktyg. Inre och yttre faktorer såsom till exempel ventilationen i testsalen spelar också roll för studenternas prestation, men också för lärarens bedömning. Messick, Kane och Bachman och Palmer har tagit fram analyseringsinstrument som skall vara till hjälp för läraren i sitt avgörande om testets giltighet. Deras modeller ifrågasätter både explicita och implicita faktorer. Modellerna är valda utifrån att de kompletterar varandra utifrån ett mikro- och ett makroperspektiv.

Min praxisorienterade undersökning utgår ifrån de nämnda modellerna och är avgränsad till att genomföras på slutrapporten som skrivs av studenterna i *Invandrarakademin*, ett högskoleförberedande program på grundläggande nivå (GRUV) och gymnasial nivå. Denna slutuppgift kan jämföras med gymnasieskolans projektarbete. I den skriftliga rapporten testas studenternas skriftliga förmåga, men också deras muntliga förmåga. Den muntliga delen testas genom att studenterna ska opponera på varandras rapporter. De ovanstående analyseringsinstrumenten kommer att användas för att avgöra rättsäkerheten i slutrapporten.

Syftet och målet med denna valideringsstudie är att denna rapports slutsatser skall leda till att den fortsatta bedömningen av dessa studenter skall bli mer rättssäker och i ett vidare perspektiv för samhället.

Konsekvenserna av bedömningen påverkar om de har möjlighet att kunna få arbete som till exempel lärare och läkare, vilket gör att vår bedömning inte bara får följder bara för studenterna själva utan även för samhället i stort. Målet är också att uppgiften skall vara utformad på ett sådant vis att invandrade akademiker skall ha samma möjligheter att bli godkända oberoende av varifrån de kommer och vilka erfarenheter som de har, vilket är helt i linje med de föreskrifter vi skall följa i vårt professionella arbete.

1.1. Syfte

Mitt syfte är att undersöka om Invandrarakademins slutrapport kan anses vara giltig och tillförlitlig som examinationsform. Jag har valt att analysera giltigheten utifrån tre olika modeller/referensramar: Messicks modell, Kanes tolkningskedja, Bachman och Palmers testkvaliteter samt språkskalan *Gemensamma europeiska referensramen för språkskalan* (GERS).

1.2. Avgränsningar

Denna uppsats är avgränsad till att endast gälla Invandrarakademins sista delkurs och resultatet kan inte appliceras på liknande kurser på samma nivå.

1.3. Disposition

Uppsatsens disposition är fördelad på så sätt att först kommer bedömningsmodeller och referensramar att presenteras. I resultatdelen kommer Invandrarakademins slutkursuppgift sättas in i dessa modeller var och en för sig. I diskussionen kommer de olika modeller, som kompletterar varandra, att diskuteras tillsammans och avslutningsvis kommer en slutsats att presenteras.

2. Bakgrund - Invandrarakademien

Högskolan i Borås anordnar en kurs för invandrade akademiker på B1 till C2 nivå enligt den *Gemensamma europeiska referensramen för språkskalan* (GERS). Kursen är på heltid och på sammanlagt 60

högskolepoäng (hp) som är förlagda över två terminer. För att få utländsk högskoleexamen validerad i Sverige krävs godkänt betyg i Sv2/SVA2 (tidigare benämnd SvB) för flertalet professioner, vilket betyder att man skall ha motsvarande kunskaper i svenska språket som en gymnasieelev som har gått en treårig högskoleförberedande-utbildning. (30 juni 2013 kommer det krävas Sv3/SVA3 och kursplanen kommer att skrivas om utefter de nya direktiven.)

Målet med kursen är att studenterna efter avslutad kurs skall få sin examen validerad och kunna söka ett arbete inom sin profession eller komplettera sin utbildning utifrån Universitets- och Högskolerådets krav för behörighet. Sv2/SVA2 krävs inom vissa yrkeskategorier för att kunna ansöka om legitimation till exempel lärare. Kursen finns med i Bedömningshandboken, vilket betyder att kursen har status av gymnasial nivå enligt Universitets- och Högskolerådet.

2.1 Invandrarakademins kursplan

Delkursen *Svenska språket i högre utbildning* har följande mål att efterfölja enligt kursplanen, vilken fastställdes 2012:

- tillämpa en för högskolestudier och/eller arbete inom yrkesområdet muntlig och skriftlig färdighet i svenska
 - skriva och strukturera en rapport, som möter den språkliga nivån C1 enligt Europarådets språkskala, som handlar om hållbar utveckling
 - självständigt och källkritiskt inhämta information
 - opponera på övriga kursdeltagares rapporter
 - reflektera kring akademisk kurslitteratur inom det egna yrkes/utbildningsområdet
 - arbeta självständigt med hjälp av handledning
- Ovanstående mål mäts mot Europarådets nivåskala (GERS).
(Högskolan i Borås 2012)

Delkursen *Svenska språket i högre utbildning*, 15 hp, går parallellt med två andra delkurser under vårterminen. De andra delkurserna är *Svenska språket i samhälls- och yrkeslivet*, 10 hp, och *Svenska språket i 1900-talslitteraturen*, 5 hp. Den sista slutrapporten som skrivs i delkursen *Svenska språket i högre utbildning* görs delvis tillsammans med samhällskunskapskursen, vilken ger behörighet i samhällskunskap A.

3. Teori och modeller

Under denna rubrik kommer tidigare forskning och litteratur inom ämnet att presenteras. Först kommer en allmän översikt över bedömning generellt och vidare följer en detaljerad presentation av de olika bedömningsmodellerna och referensramarna som jag har valt att använda för min analys.

3.1 Bedömning

Bedömning kan ses ur två olika perspektiv. Det traditionella är att en student efter avslutat område får visa läraren genom ett prov, vilket vanligtvis brukar vara skriftligt eller muntligt, ifall denne har förvärvat de kunskaper som är uppsatta i kursmålen. Provet bedöms sedan utifrån mer eller mindre uttalade kriterier och inte sällan görs detta genom ett poängsystem. Det numeriska värdet/summan avgör om eleven anses vara godkänd eller inte. Detta är en summativ bedömning. Formativ bedömning görs ur ett annat perspektiv där det är processen som är det centrala istället för att endast kontrollera slutprodukten. En annan benämning för formativ bedömning är bedömning för lärande (BFL). William (2013) och Lundahl (2011) är de stora förespråkarna för formativ bedömning. Formativ och summativ bedömning skall inte ses utifrån att den ena är bättre än den andra utan att de kan komplettera varandra

Det finns flera faktorer som gör att man kan dra mer eller mindre korrekta slutsatser utifrån bedömning och detta oberoende av om det är en formativ eller en summativ bedömning som genomförs. Det skall dock påpekas att i test generaliserar man en individs språkförmåga utifrån vad hon har visat på testet, och bedömer vad man tror att hon skulle klara av i andra sammanhang (Sundberg 2007: 78).

Bachman och Palmer (2010) skriver att man bör konstruera uppgifter utifrån en viss domän, för att testet skall bli relevant och för att man skall kunna fatta rätt beslut utifrån testet. I ett högskoleförberedande test bör det finnas uppgifter som ingår i denna domän såsom till exempel att skriva rapporter och reflektionsuppgifter dvs sådant som är vanligt förekommande i denna domän. Detta betyder att testkonstruktören måste besitta kunskaper om vad som utmärker språket i domänen som skall testas och bedömas, ifall kunskaperna saknas är det troligt att tolkningarna av testresultatet blir meningslösa.

En annan faktor som påverkar testets giltighet är att *konstrukten*, alltså hur en uppgift är utformad för att få önskad output, skall bygga på att en och samma språksyn följs genom hela testet (Fulcher & Davidsson 2007:36f) för att korrekt bedömning skall kunna göras. Slutsatserna man drar av bedömningen blir oanvändbara om man i konstrukten endast gör uppgifter som testar vardagsspråk, men skall mäta formellt skriftspråk.

En annan faktor som påverkar korrektheten av bedömningen är att studenter underpresterar ifall testet är för lätt, vilket medför att studentens språkförmåga inte kommer i sin rätta dager. Detsamma gäller ifall språktestet saknar relevans till exempel ifall en testdeltagare ombeds beskriva hur man lagar en motor på en bil och testdeltagaren aldrig har genomfört denna syssla. Testet saknar relevans för testdeltagaren och då är det inte möjligt att bedöma testdeltagarens språkliga nivå med hjälp av detta test.

En annan sak som gör testresultat meningslösa är ifall generaliserbarheten uteblir, vilket gör test missvisande enligt Lundahl (2011:73) som skriver att det är viktigt att inte övertolka resultat. Han använder de nationella proven som ett konkret exempel. De nationella proven mäter bara delar av lärmålen, vilket betyder att en individs hela språkliga förmåga inte kan bedömas utifrån ett skriftligt prov.

Baker menar att testbedömare bedömer olika på grund av olika faktorer. En av faktorerna som hon nämner är att vissa bedömaren ger studenter ett högre betyg eller godkänt ifall man vet att bedömningen får stora konsekvenser för studenten.

I felt like giving this a 1, but because I knew the effects of that decision on the student, I was more careful/lenient in accepting that all the tasks (esp. Writing parents) were included. (Baker 2012, appendix B)

Reliabiliteten minskar om olika bedömare bedömer samma test (Fulcher & Davidson 2007:30) och detta medför att bedömningen inte blir rättssäker gentemot testdeltagarna.

3.2 Gemensam europeisk referensram för språk (GERS)

Common European Framework of Reference (CEFR), också kallad *Gemensam europeisk referensram för språk (GERS)*, är en modell som utarbetades för att européer skulle kunna självskatta sina

språkkunskaper i främmande språk och att olika språkliga nivåer skulle kunna vara jämförbara i alla europeiska länder.

GERS tillkom mellan åren 1989-1996 av Europarådet för att det fanns ett behov av en gemensam referensram för personer som arbetade med moderna språk i Europa. GERS skulle användas vid utarbetande av läromedel, tester, kursplaner och läroplaner (Skolverket 2009:1). Referensramen skulle alltså underlätta för till exempel lärare i bedömningen för individers uppnådda nivå i olika moderna språk.

Syftet med *Gemensam europeisk referensram för språk* är att övervinna de kommunikationshinder mellan människor som arbetar med moderna språk, som uppstår till följd av de olika utbildningssystemen i Europa. (Skolverket 2009:1)

För att GERS skulle bli användbart i så många europeiska länder som möjligt bortsåg man från kulturella värderingar och det kulturbundna språket. Det som betonas är språkhandlingarna (McNamara & Roeber 2006). Detta medför att språkanpassningen till plats, ämne och person inte är tydlig. EU har arbetat för högre mobilitet och en gemensam skala för bedömning av språkkunskaper ansågs underlätta om flertalet använder sig av samma mall för bedömning. Gemene man använder sig av ord som till exempel nybörjarnivå, men detta ord är tolkningsbart. GERS beskriver vilka språkliga kompetenser man besitter i målspråket när man befinner sig på en viss nivå.

Referensramen kan utgöra ett hjälpmedel för ansvariga inom utbildningssektorn, kursutvecklare, lärare, lärarutbildare, examinationsinstitutioner etc. när de reflekterar över sina rutiner för att fastställa och samordna sitt arbete samt att säkerställa att de möter de verkliga behoven hos de studerande som de har ansvar för. (Skolverket 2009:1)

GERS bygger på en individuell funktionell språksyn. Det betyder att modellen beskriver och ger exempel på vilka funktioner i vardagen som vi kan behärska i det testade språket och bortser ifrån ifall korrekt grammatik används eller ej. GERS är såsom anges i citatet ovan inte avsedd för att användas som bedömningsmall för språktester. Van Avermaet (2009:34) är mycket kritisk till användningen av GERS idag och menar att GERS utformades för bedömning av främmande språk och inte för undervisning eller bedömning av andraspråk. GERS förutsätter literacitet och riktar sig därmed endast till dem med en god skolbakgrund. I *Gemensam referensram för språk* (2007) beskrivs att

GERS -skalan skall användas för självskattning. ”Dessutom skall alla användare kunna beskriva sina egna mål etc. utifrån referensramen” (Skolverket 2009:7), vilket visar att GERS-skalan skall användas formativt. Istället har GERS -skalan kommit att användas i bedömningen av summativa tester.

GERS genomsyras av en funktionell språksyn. Det betyder att språket används som en funktion i kommunikationen människor emellan och att språket skapas gemensamt. Vardagsdomänen och pragmatiken är i fokus i GERS och att språk är ett kommunikationsmedel mellan minst två aktörer i samhället.

GERS är i teorin uppdelad i läsförståelse, skriftlig produktion, muntlig produktion och hörförståelse, men i praktiken är inte dessa delar skilda utan dessa delar tillhör en helhet. Skaparna av modellen har en medvetenhet om det och säger följande: ”Detta ställer oss inför psykologiska och pedagogiska problem eftersom kommunikation är något som tar hela människan i anspråk” (Skolverket 2009:1). Det är också tydligt att det är nödvändigt att klassificera språket i separata komponenter för att det skall bli hanterbart.

Nedan följer kriterierna på C1-nivå enligt GERS:

Skriftlig förmåga - Jag kan uttrycka mig i klar och väl strukturerad text med utförligt angivna synpunkter och förklaringar. Jag kan skriva om komplicerade förhållanden i ett brev, en uppsats eller en rapport och argumentera för vad jag anser är viktigt. Jag kan välja en stil som är anpassad till den tänkte läsaren.

Tala - Jag kan uttrycka mig flytande och spontant utan att alltför tydligt söka efter rätt uttryck. Jag kan använda språket flexibelt och effektivt för sociala, intresse- och yrkesanknutna ändamål. Jag kan formulera idéer och åsikter med viss precision samt med viss skicklighet anpassa mig efter den person jag talar med. Jag kan klart och detaljerat beskriva komplicerade ämnesområden med sidoteman och anknytningar. Jag kan utveckla speciella aspekter samt runda av framställningen med en konsekvent slutsats.

Förståelse - Jag kan förstå långa och komplicerade faktatexter liksom litterära texter och jag uppfattar skillnader i stil. Jag kan förstå fackartiklar och längre instruktioner även inom områden som ligger utanför mina intressen och erfarenheter. (Skolverket 2013)

3.3 Kanes tolkningskedja

Kane har utvecklat en modell för ett systematiskt arbete med tillförlitliga tolkningar och bedömningar av tester som visar på hur testresultat och tolkning hör ihop. Modellen kallas *Chain of inferences* och består av fem olika kedjor vilka är sammanlänkande och för en rättssäker och objektiv bedömning skall det inte finnas några brister i kedjan. De olika kedjorna i Kanes tolkningskedja heter:

1. *Eliciting student performance: Task design*
1. *Scoring the performance: Scoring procedures.*
2. *How typical is this score? Reliability, generalizability*
3. *What does the score mean? Interpreting*
4. *Using the score to make decisions: Consequences* (Kane 1992).

Utifrån *Chain of inferences* kan man konkret se hur ett test bör hänga ihop hela vägen från testkonstrukt till de konsekvenser det får utifrån ett mikro- och makroperspektiv.

Den första kedjan för att nå det slutgiltiga målet är testkonstruktionen - *Eliciting student performance: Task design*. Denna kedja är utformandet av tester. I utformningen av tester bör testkvaliteterna ta i beaktande till exempel frågor som: Hur lång tid får testdeltagaren ta i anspråk? Hur formuleras instruktionerna? Är instruktionerna muntliga eller skriftliga? (Bachman & Palmer 2010:66ff).

Nästa kedja är rättning - *Scoring the performance: Scoring procedures*. I denna kedja bör det finnas en bedömningsmatris och kurskriterier som utgår ifrån de uttalade kursmålen. Inom språklig bedömning finns en uppsjö av bedömningsmodeller till exempel felanalys, den pragmatiska trappan, processbarhetsanalys, performansanalys, genreanalys SFL och språkhandlingsteori. Beroende på vilken av dessa nämnda som används kan resultatet bli varierat t.ex så fokuseras bedömningen på lingvistiska fel i felanalys medan performansanalysen fokuserar på vad studenten kan bemästra lingvistiskt.

Den tredje kedjan är bedömningens generaliserbarhet - *How typical is this score? Reliability, generalizability* och den handlar om hur tillförlitligt resultatet av bedömningen är. Baker (2012) påvisar att det finns socio-kognitiva skillnader mellan olika testbedömare som kan förklara skillnaderna i deras bedömningar av samma resultat. Baker har i sin forskning antecknat testbedömarnas kommentarer när de har bedömt en skriftlig text och där ser man tydligt att testbedömarna bedömer

utifrån olika variabler. En del bedömde utifrån olika slags känslor till exempel de konsekvenser bedömningen får för testtagaren medan andra bedömde utifrån till exempel ett lexikalt, strukturellt eller lingvistiskt perspektiv. Baker kommer också fram till att testbedömare har olika strategier för bedömning. En del går efter principen att första intrycket varar. Dessa bedömare bedömer texten efter några få meningar och sedan letar de efter motiveringar i resten av texten för att bibehålla den första uppfattningen av texten. Andra behöver läsa texten igenom helt och hållet både två och tre gånger. Tillförlitligheten ökar ju konkretare matriser som finns för bedömningen (Baker 2012:225-248).

Nästa del i kedjan är tolkning av resultatet - *What does the score mean? Interpreting*. Vad är testets relevans och vad är nyttan av testet? Till exempel skulle det varit relevant att testa ifall studenterna har kunskaper i hur man skriver en rapport ifall de går en högskoleförberedande utbildning och att testa ifall de har tillräckligt med ord på målspråket för att kunna slutföra denna uppgift.

Den sista delen i kedjan är konsekvenser - *Using the score to make decisions: Consequences*. Vilka konsekvenser får testet för individen, för gruppen och för samhället generellt? Huang anser att test kan vara motiverande för studenterna, men att detta inte alltid är fallet.

A test or a required assignment, if not carefully designed, can easily demotivate and help in shaping learners' helplessness beliefs, even though the original intention was perfectly justifiable. (Huang 2012:75)

Alla delarna i kedjan hänger ihop eftersom de olika delarna påverkas av varandra. Om man liknar kedjan vid bryggor, så kan man dra slutsatsen att ju söndrigare bryggorna är i början desto svårare är det att nå torrskodd hela vägen ut till den sista.

3.4 Messicks giltighetsmodell

Messick (1989) har utformat en modell där ett test skall ses ur olika perspektiv och då framför allt utifrån en social dimension. Språket finns alltid i en social kontext och då är det relevant att ta reda på vilka sociala normer och värden som finns i ett test. Vissa forskare är kritiska till om begreppet validitet verkligen kan kopplas till sociala värden och dess konsekvenser, eftersom det delvis handlar om subjektiva iakttagelser. Messick menar att det är just subjektiviteten och det implicita vi behöver lyfta fram för att komma åt validiteten och se de orättvisor som finns inbäddade i test. Messick utgår ifrån att validitet hör samman med

rättvisa. Konsekvenserna som individerna får av ett godkännande eller ett underkännande skall också finnas i beaktande liksom vilka konsekvenser bedömningen av testet får för samhället (McNamara & Roever 2006).

I figur 1 presenteras Messicks giltighetsmodell genom hans rutmodell. Konstrukten är basen och med konstrukt menas här hur en uppgift är utformad för att kunna påvisa specifika kunskaper.

Messicks giltighetsmodell	Testtolkning (Vad vill man mäta med rapporten?)	Testgenomförande/användande (Vad mätte rapporten/Vad är nyttan?)
Bevisbas (bevis för att stödja påståenden)	Ruta 1: Vad är konstrukten? Vad testas och vad testas inte?	Ruta 2: Vad är nyttan och vad är relevansen? Är tolkningarna meningsfulla, användbara och rättvisa i en given kontext?
Konsekvensbas (Den sociala kontexten i testandet)	Ruta 3: Vilka implicita och explicita sociala och kulturella värderingar testas?	Ruta 4: Vilka konsekvenser får resultatet för studenten och för samhället?

Figur 1: Messicks modell (efter McNamara & Roever 2006:13)

Ruta ett i Messicks modell handlar om vad det är som skall mätas/bedömas och vad som inte skall bedömas i testet och detta hänger också ihop med den konkreta utformningen av uppgiften liksom testkvaliteterna. Svaren på alla dessa frågor påverkar hur testet skall utformas. Denna ruta är jämförbar med den första kedjan i Kaness tolkningskedja (se avsnitt 3.3).

Messicks andra ruta kan paras ihop med Kaness fjärde tolkningskedja. Vilken verklighetsförankring/nytta har testet?

Den tredje rutan analyserar de explicit/implicit sociala värden som testas och om det finns sociala värden som testas explicit/implicit?

Ruta nummer fyra är samma som den sista delen i Kaness tolkningskedja. Testet skall ses ur ett konsekvensperspektiv.

Det finns stora likheter mellan Messicks giltighetsmodell och Kaness tolkningskedja, men den stora skillnaden dem emellan är den sociala dimensionen som finns med i Messicks modell. Ett tests giltighet måste ses utifrån de sociala värden som testas implicit, anser Messick, medan Kane poängterar examinatorernas bedömning och bedömningens värde för testets tillförlitlighet.

3.5 Bachman och Palmers testkvalitéer

När ett test genomförs finns det många faktorer som kan påverka resultatet och testets giltighet. Bachman och Palmer (2010) tar i beaktande de fysiska och de yttre faktorerna i sin modell till skillnad från Kane och Messick som till största del har psykiska och inre faktorer i fokus. Bachman och Palmers olika områden är:

1. *Characteristics of the setting*
2. *Characteristics of the rubric*
3. *Characteristics of the the input*
4. *Characteristics of the expected response*
5. *Relationship between input and expected response (Bachman & Palmer 2010)*

Dessa olika förutsättningar ger upphov till frågor som man som testgenomförare och testkonstruktör skall fråga sig själv innan testets genomförande. Nedan finns några frågor som skall användas för att ifrågasätta testets giltighet och trovärdighet. Svaren på frågorna avgör påverkansfaktorn på den slutgiltiga bedömningen som görs av testet.

Förutsättningar och tillgänglighet (*Characteristics of the setting*) kan påverka testets giltighet till exempel så kan validiteten påverkas om examinationen skall utföras med hjälp av IT-verktyg. Ifall vissa av studenterna har otillräckliga IT-kunskaper för att genomföra testet blir den summativa bedömningen av testet missvisande.

Testets utformning (*Characteristic of the rubric*) är en annan konkret faktor som påverkar slutresultatet. Språknivån i instruktionerna till examinationen måste vara i nivå med nivån som skall testas eller under. Tiden är en annan faktor. Denna måste vara utprövad så att det finns en rimlig chans för studenterna att klara testet inom utsatt tid.

Inputen (*Characteristics of the input*) som studenterna får i form av till exempel bilder, texter och ljud behöver testtagaren förstå för att kunna genomföra testet korrekt och den förväntade outputen (*Characteristics of the expected response*) är examinatorns förväntningar på studenternas produktion. Studenten måste förstå vad som förväntas. Relationen mellan de två ovannämnda, input och output (*Relationship between input and expected response*) påverkar testets slutresultat liksom examinationens giltighet. Studenterna måste till exempel ha fått en chans att tillgodogöra sig materialet som de skall bedömas på (Bachman och Palmer 2010).

4. Metod och material

Detta kapitel beskriver detaljerat vilka kvalitativa metoder som har använts för denna uppsats. Först genomfördes en valideringsstudie i form av genomgång av material som finns tillgänglig för studenterna i Invandrarakademin och efter denna materialgenomgång följde två djupintervjuer, vilka kommer beskrivas nedan tillsammans med valideringsstudien.

4.1 Material och metod

Den kvalitativa metoden som har använts i uppsatsen genomfördes i form av en valideringsstudie med utgångspunkt från språkliga bedömningsmodeller och språkliga referensramar tillsammans med två kompletterande djupintervjuer. Underlaget som har använts i valideringsstudien är den giltiga kursplanen för Invandrarakademin, skriftliga instruktioner som finns till förfogande för studenterna på lärplattformen och inspelade lektioner i rapportskrivning.

Instruktionerna som ges skriftligen och muntligen till rapportskrivningen och kursplanen för Invandrarakademin analyserades först utifrån Messicks modell därefter i Kanes tolkningskedja och Bachman och Palmers testkvalitéer. Sist analyserades det material som fanns till förfogande utifrån GERS. Syftet besvarades utifrån de olika perspektiv som dessa har.

Valideringsstudien kompletterades med intervjuer med de två examinatorerna på kursen. Dessa två personer är också kursansvariga, testkonstruktörer och undervisande lärare i kursen och valdes på grund av att de är ansvariga för fallstudiens examination. Intervjuerna genomfördes i slutet av studien för att komplettera det övriga materialet. Fokus i intervjuerna var på de frågor i analyseringsinstrumenten som inte hade blivit besvarade genom det skriftliga materialet. Intervjuerna var semistrukturerade och har spelats in på en mp3-spelare för att jag som intervjuare skulle kunna fokusera på samtalet. Intervjufrågorna finns i sin helhet i bilaga ett i resultatet kommer svaren som framkom vid intervjuerna presenteras tillsammans med det övriga materialet. De två intervjuade personerna kommer i resultatdelen benämnas informanter.

4.2 Etiska principer

Vetenskapsrådets forskningsetiska principer (2005) har efterföljts angående information, samtycke, konfidentialitet och nyttjande av information. De intervjuades identiteter är fingerade och deltagandet var frivilligt. De intervjuade blev informerade om detta muntligen liksom att de har möjlighet att läsa rapporten för godkännande innan publicering.

Under analyserandet av intervjuerna hade jag i beaktande Björkdahl Ordells ord "...att från början beakta om de personer som involveras i undersökningen på något sätt utlämnas och kan komma till skada" (Björkdahl Ordell 2007:21).

4.3 Giltighet och tillförlitlighet

"Forskningens trovärdighet handlar om relationen mellan de tolkningar som forskaren gör och de data som ligger som underlag för tolkningarna" (Eklund 2011:19). Det är därför högst väsentligt att diskutera rapportens giltighet och tillförlitlighet.

Giltighet innebär att man undersöker det man har för avsikt att undersöka. Syftet med rapporten är ta reda på om rapportskrivningen kan anses vara giltig och tillförlitlig som examinationsform i svenska som andraspråk i Invandrarakademin, som är en högskoleförberedande kurs, vid Högskolan i Borås.

Tillförlitlighet är att en mätning är tillförlitlig, stabil och inte utsatt för några slumpinflytelser (Trost 1994). Tillförlitligheten blir lägre i kvalitativa undersökningar än i kvantitativa undersökningar. Man kan särskilja fyra komponenter i det sammansatta begreppet reliabilitet: kongruens, precision, objektivitet och konstans. För forskningens tillförlitlighet är det av största betydelse att frågorna som ställs är de samma i alla intervjusituationer, vilket innebär kongruens. Informanterna fick samma frågor av mig. Precision är att man som intervjuare tolkar svaren som man får vid intervjuandet på liknande sätt i olika intervjusituationer. Intervjuerna spelades in och analysen genomfördes efter att de två intervjuerna var slutförda. Precisionen är hög vid enkäter då det är lätt att registrera vilken ruta som har markerats av den som fyllt i enkäten, men blir lägre vid intervjuer. Objektiviteten är hög eftersom det endast var en intervjuare som utförde intervjuerna. Människors inställningar och attityder är föränderliga eftersom människors attityd kan hinna ändras på grund av inflytande från diverse faktorer som man inte kan bortse ifrån under denna tid. Konstansen är

hög eftersom intervjuerna genomfördes under samma vecka för denna uppsats (Trost 1994:58).

Jag anser att rapporten håller en relativt hög tillförlitlighet och giltighet utifrån ovanstående parametrar.

5. Resultat

I resultatdelen kommer det analyserade materialet att presenteras utifrån de olika modellerna och referensramarna. Valideringsresultatet kommer att presenteras tillsammans med resultatet från de två intervjuerna som har genomförts med de kursansvariga för kursen och som också är examinatorer och testkonstruktörer. Analys och diskussion kommer att följa i nästa kapitel. Resultaten kommer att presenteras enligt följande ordning; Kanes tolkningskedja, Messick, GERS och Bachman och Palmer testkvalitéer.

5.1 Rättssäker bedömning

Lärarna på Invandrarakademin har valt att examinera studenterna i Sv2 genom att studenterna skall skriva en akademisk rapport som har koppling till studenternas profession och dessutom till hållbar utveckling. Motiveringen till varför de inte har valt att använda *Test i svenska för universitets- och högskolestudier* (TISUS), som är ett behörighetsgivande test på samma nivå och som är utprovat av Institutionen för svenska och flerspråkighet på Stockholms Universitet, är att TISUS anses vara för brett i sin testning av svenska språket. Dessutom anser lärarna att de är behöriga att examinera Sv2 och att man då kan göra detta lokalt och inte köpa in denna tjänst utifrån.

En av informanterna definierar rättssäker bedömning enligt följande: ”Att den inte går att påverka genom påtryckning utifrån och att flera lärare som sambedömer ger samma bedömning.” Vidare säger denne att det finns press från studenterna sida att de skall bli godkända inom ramen för terminen och detta för att studenterna vill kunna påbörja fortsatta studier. ”En del studenter tror de skall bli godkända bara för att de har lämnat in en uppgift” (Informant 1).

Alla rapporterna läses av de båda examinatorerna och efter genomläsningen diskuterar de rapporternas fördelar och nackdelar.

Examinatorerna beslutar tillsammans vad som behöver kompletteras ifall några kompletteringar behöver göras.

Ibland behöver vissa studenter ytterligare en termin för att kunna bli godkända. Jag tror redan nu att vi har några studenter som inte kommer bli godkända nu i augusti t.ex. x och y. (Informant 1)

5.2 Kanes tolkningskedja

Resultaten som går att analysera enligt Kanes tolkningskedja har satts in i Kanes fem olika kedjor som alla påverkar varandra och slutligen testets giltighet.

1. *Eliciting student performance: Task design.* (McNamara & Roever 2006:26) Instruktionerna för rapporten ges både skriftligt och muntligt där de skriftliga instruktionerna finns till förfogande att ladda ner från lärplattformen och föreläsningar som har rört rapportskrivning finns att streama. Digitala presentationer som har använts vid föreläsningarna finns att ladda ner via lärplattformen. Studenterna har två månader på sig för att skriva rapporten. Delkursen *Svenska språket i högre utbildning* (15 hp) är förlagd under hela vårterminen och två andra halvfartskurser (10 + 5 hp) pågår parallellt med rapportskrivningen. De skriftliga formuleringarna i instruktionerna ligger på en hög nivå i svenska språket. Instruktionerna är inte medvetet skrivna på denna nivå, säger informant 2, men menar å andra sidan att instruktionerna blir omedvetet en del av bedömningen eftersom ifall inte studenterna förstår instruktionerna kan de inte fullfölja uppgiften korrekt enligt mallen.

2. *Scoring the performance: Scoring procedures.* ”Vi utgår ifrån GERS i den språkliga bedömningen, men jag bedömer mest enligt erfarenhet” (Informant 2). Vidare säger informanten att en mer detaljerad bedömningsmatris för den språkliga förmågan hade riskerat att göra bedömningen för summativ.

3. *How typical is this score? Reliability, generalizability*

Självständiga arbeten i form av rapportskrivning är en vanligt förekommande examinationsform vid universitet och högskolor, anser informant 1, och ifall studenterna kan uppvisa kunskaper i att de kan referera och hantera språket på lämplig nivå så anser denne att resultatet är tillförlitligt.

4. *What does the score mean? Interpreting.* Testets relevans är att examinationen testar den skriftliga förmågan inom högskoledomänen och till viss del även den muntliga färdigheten genom att opponering genomförs. Nyttan av testet är att studenten tränas på målspråket i

formell svenska och kan uppvisa kunskaper inom rapportskrivning som är en stor del inom högskoledomänen.

5. *Using the score to make decisions: Consequences.* Konsekvenser av testet för individen är att de har möjlighet att komplettera sina tidigare studier vid ett universitet eller en högskola i Sverige efter ett godkänt betyg. Studenter med godkända utbildningar har möjlighet att söka arbete inom sin profession. Konsekvenserna vid underkänt är att studenterna måste vänta med sina kompletterande studier, vilket får ekonomiska effekter för individen och i en del fall för samhället. Inom många professioner behövs inte Svenska 2-nivå för att ett fullgott arbete skall kunna utföras och i dessa fall missar svenska samhället god arbetskraft, enligt en av informanterna.

5.3 Messicks giltighetsmodell

Messicks giltighetsmodell består av fyra rutor där testtolkning och testgenomförande är två parametrar som ställs mot bevisbasen och konsekvensbasen. Analysresultaten kommer att presenteras ruta för ruta i figur 2.

Messicks giltighets Modell	Testtolkning (Vad vill man mäta med rapporten?)	Testgenomförande/användande (Vad mätte rapporten/Vad är nyttan?)
Bevisbas (bevis för att stödja påståenden)	Ruta 1: Konstrukten är att skriva ett självständigt arbete enligt gängse regler för akademiskt skrivande för att bevisa att man har tillräckliga språkkunskaper för universitetsstudier på målspråket.	Ruta 2: Uppgiften mäter ifall studenten kan följa en mall och att följa skriftliga och delvis muntliga instruktioner. Referensskrivande, sammanfattningar och reflekterande texter testas också tillsammans med källkritik och att muntligen kunna framföra konstruktiv kritik.
Konsekvensbas (Den sociala kontexten i testandet)	Ruta 3: Studenterna testas ifall de kan skriva enligt den svenska modellen för rapportskrivning liksom om de kan uttrycka sina egna tankar och åsikter i tal och skrift. Andra sociala värden som testas är att vara källkritisk.	Ruta 4: Konsekvensen för individen är möjligheten till fortsatta studier eller arbete. Konsekvensen för samhället är möjligheten till arbetskraft eller förlust av god arbetskraft. Bättre kunskaper i formell skriftlig svenska?

Figur 2: Invandrarakademiens slutuppgift i Messicks modell

Ruta 1: Konstrukten är enligt en av informanterna att studenten skall kunna skriva en rapport och kunna opponera på övriga kursdeltagares rapporter för att kunna påvisa att de har tillräckliga kunskaper i svenska språket för att kunna studera vidare och att de uppvisar att de kan skriva ett självständigt arbete efter givna akademiska ramar. Den språkliga nivån skall vara C1 enligt GERS. Med denna konstrukt vill man testa studenternas förmåga att kunna uttrycka sig skriftligt inom genren rapportskrivning och även testa studenternas förmåga att ge konstruktiv kritik på andra studenters rapporter.

Studenternas skriftliga förmåga inom andra genrer testas inte. Giltigheten för denna konstrukt är att den likställs med gymnasiets projektarbete liksom att rapportskrivandet är den genre som ligger inom högskoledomänen.

Ruta 2: Kursen är på högskoleförberedande nivå alltså på gymnasienivå och på de teoretiska gymnasieprogrammen genomför eleverna ett projektarbete. Relevansen av uppgiften är att den kan likställas med detta projektarbete. Nyttan med examinationsuppgiften är att kunna visa att man kan ”tillämpa en för högskolestudier och/eller arbete inom yrkesområdet muntlig och skriftlig färdighet i svenska” (Högskolan i Borås 2012).

Vi vill ge studenterna verktyg som förbereder dem för kompletterande studier på svenska inom deras profession. Vi har försökt att låta dem göra en typisk rapport för A och B-nivå där fokus ligger på att kunna hantera språket, formatet och att kunna referera. (Informant 1)

Ruta 3: De kulturella och sociala värderingar som testas i delkursen är om studenten kan ”tillämpa en för högskolestudier och/eller arbete inom yrkesområdet muntlig och skriftlig färdighet i svenska... strukturera en rapport...självständigt och källkritiskt inhämta information...opponera på övriga kursdeltagares rapporter...reflektera kring akademisk kurslitteratur inom det egna yrkes/utbildningsområdet...arbeta självständigt med hjälp av handledning” (Högskolan i Borås 2012).

Ruta 4: Konsekvenserna för den individuella studenten av ett godkänt på delkursen är att bli behörig till högskole- och universitetsstudier. För vissa yrkeskategorier är konsekvensen av ett godkänt att kunna ha möjlighet till att kunna ansöka om legitimation inom sin profession till exempel för lärare. Följderna av ett underkänt är enligt en av de kursansvariga:

Om de inte blir godkända kan samhället förlora duktiga yrkesmän som skulle kunna lära sig svenska språket i det praktiska arbetet å andra sidan skulle ett godkännande av läkare kunna få katastrofala följder. (Informant 2)

5.4 Backman och Palmers testkvaliteter

1. Förutsättningar/tillgänglighet (*Characteristics of the setting*)

Förutsättningarna för studenterna är att de har två månader på sig från synopsis till färdig rapport under delkursen. En annan delkurs i 1900-tals litteratur pågår parallellt. Under denna tid så har de två inbokade handledningstider, men handledare finns till förfogande under hela tiden via mail, men det finns också möjlighet att boka besök. Datorsalar på högskolan finns till förfogande från kl. 6.00–21.00 och högskolebiblioteket är öppet kl. 8.00–20.00 (något mindre öppettider under helger). Bibliotekarierna finns tillgängliga under hela denna tid för frågor. Alla instruktionerna för rapporten finns att tillgå på lärplattformen. Föreläsningar som tar upp akademiskt skrivande är filmade och är möjliga att streama. IT-lektioner ges inte och alla studenterna har olika grader av IT-kunskaper beroende på vilket land de kommer ifrån och är också beroende på profession och tidigare arbetserfarenhet.

Enligt en av de intervjuade informanterna kan man inte tala om relationen mellan lärare och student i generella termer utan det är individuellt. Vidare säger informanten att vissa studenter gillas bättre av en del lärare och andra studenter gillas av andra lärare och vice versa och detta kan inte helt bortses ifrån. Många av studenternas examinationer görs anonymt med tanke på detta, men slutrapporten kan inte lämnas in anonymt eftersom handledare och examinator är samma person. Relationen med bibliotekarierna är en annan faktor som kan påverka slutresultatet och på denna fråga svarar informanten:

Studenterna är rädda för att be om hjälp generellt och några vet inte om att de kan få hjälp från bibliotekarierna. När jag har hänvisat studenter till biblioteket har jag fått fågelholksreaktioner. (Informant 1)

2. Testets utformning (*Characteristic of the rubric*)

Nivån i språket är hög, till exempel används följande formuleringar: *lyft frågor, rapportens förtjänster, väva ihop och ostrukturerad hopning*. Instruktionerna är inte medvetet någon del av testet, enligt en av de kursansvariga, men de blir detta indirekt. ”Ifall studenterna inte förstår

instruktionerna så klarar sig de inte” (Informant 2). Rapporterna är inte kodade, vilket betyder att de inte lämnas in anonymt.

3. Input (*Characteristics of the input*)

Den stora inputen, enligt en av informanterna, är föreläsningarna, seminarierna och handledningstillfällena. Vidare påpekar denne att ”Ett stort problem är att vi inte har varit tillräckligt tydliga med hur viktig närvaron är för att kunna få ett bra resultat” (Informant 1). Material som skall finnas som underlag till rapporten får studenterna själva leta upp. Studenten kan få hjälp i sitt sökande av handledaren och/eller högskolebibliotekarier. ”Studenterna har inte fått en specifik lektion om källkritik, men de har fått uppgifter med källkritiska inslag” (Informant 2) ”Ett mänskligt skelett för att illustrera en struktur används i kursen, men hur tolkas/förstås denna av rapportskrivarna det kan jag bara spekulera kring” (Informant 1). Vidare säger informanten att det inte finns någon medveten tanke med valet av bild eller text, utan informanten har inte reflekterat över detta innan frågan ficks. Studenterna skall utgå ifrån en mall i sitt skrivande av rapporten. Denna mall har använts i en tidigare delkurs på GRUV-nivå, så studenterna är något familjära med denna mall.

4. Förväntad output (*Characteristics of the expected response*)

Den förväntade outputen är, enligt en av informanterna, att studenterna kan skriva en akademisk rapport som är språkligt acceptabel utan störande fel och där källor används korrekt.

5. Relationen mellan input och output (*Relationship between input and expected response*) Inputen påverkar outputen, anser en av informanterna.

Det är stor skillnad på de studenter som har tillgodosett sig materialet mot de som inte varit närvarande på lektionerna och inte heller laddat ner materialet från Pingpong. (Informant 1)

Vidare säger informanten att inputen är något som de måste tänka igenom bättre till nästa kursstart.

5.5 Gemensam europeisk referensram för språk (GERS)

Gemensam europeisk referensram för språk (GERS) används som bedömningsmatris av den skriftliga förmågan. Bedömningen skall inte vara godtycklig anser en av de kursansvariga utan bygger på vad som står i GERS. ”Vi utgår ifrån GERS i den språkliga bedömningen, men jag bedömer mest enligt erfarenhet” (Informant 1).

GERS är vald utifrån att

“GERS anger en språklig nivå som är generell och inte lika inriktad mot skola som gymnasiebetyg är. GERS handlar om kommunikationsförmåga i relation till andra människor i samhället.”
(Informant 1)

6. Diskussion och slutsats

I detta kapitel kommer först resultatet att diskuteras utifrån teoridelen och den vetenskapliga förankringen. Denna del följs av en metoddiskussion där de valda metoderna för denna uppsats diskuteras. Diskussionsdelen avslutas med uppsatsens slutsats.

6.1 Resultatdiskussion

Vad är egentligen en rättssäker bedömning och kan en bedömning egentligen vara hundra procent rättssäker? En rättssäker bedömning skall vara saklig, opartisk liksom att allas likhet skall beaktas. De modeller som jag har använt för att analysera vad som egentligen bedöms i Invandrarakademins slutrapport medvetandegör komplexiteten i testkonstruktion och i den efterföljande bedömningen. Det kan på ytan verka som ett perfekt prov, men när det ses ur flera olika perspektiv så upptäcks komplexiteten. I bedömandet av svenska som andraspråksstudenter tillkommer ytterligare en dimension som måste tas hänsyn till och det är den kulturella aspekten.

Messick tar upp den kulturella dimensionen som en variabel som skall beaktas i konstruktionen av ett test/och eller i avgörandet av ett provs giltighet. GERS (Skolverket 2009) har däremot medvetet bortsett från kulturella värderingar i referensramen, eftersom den är tänkt att användas interkulturellt. Men går det att eliminera den kulturella dimensionen? Det går inte enligt Messick, eftersom språket alltid finns i en social kontext (1989). I akademiskt skrivande i Sverige på grundnivå, vilken Invandrarakademins studenter skall visa kunskaper på att de behärskar, är fokus på reflekterande, analyserande och diskuterande samt att de ska vara källkritiska. Många av studenterna i kursen kommer

dock från länder där dessa egenskaper nedvärderas och där memorering istället är i fokus. Dessa studenter är i många fall vana vid att läsa in långa texter och sedan svara på kortare eller längre frågor och vid skrivandet av en akademisk rapport misslyckas dessa studenter på grund av de kulturella akademiska skillnaderna. För dessa studenter blir uppgiften inte en examination i svenska språket utan i att förstå en viss typ av genre där det prioriteras att man kan uttrycka sina egna tankar i skrift.

Uppgiften som har valts att använda som slutuppgift för att kunna uppvisa att studenterna har nått gymnasial nivå i svenska som andraspråk är ändå relevant, anser jag, även med kännedom om de värden som implicit examineras. Rapportskrivning är en vanligt förekommande genre inom universitetet, vilken är den gemensamma domänen för studenterna i denna kurs. Rapportskrivning är också en genre som används som slutuppgift vid kommunala gymnasieskolor och eftersom denna kurs likställs med dessa får de kursansvariga medhåll från mig att uppgiften är relevant. Ämnet hållbar utveckling är högaktuellt inom alla professioner och är också ett brett ämne som kan vinklas och ses ur olika perspektiv beroende på vilken profession och kulturell bakgrund som studenten har. Hållbar utveckling är ett ämne som skall genomsyra all verksamhet vid Högskolan i Borås och det anses därför relevant att detta ämne har beslutats vara utgångspunkt för studenternas skrivande eftersom kursen är högskoleförberedande.

Studenterna som studerar på Invandrarakademin kommer från olika länder och har olika professioner. IT-undervisningen ser olika ut i olika länder. Vissa studenterna har IT som en del av sin profession såsom systemutvecklare medan andra aldrig har använt dator tidigare och detta medför naturligtvis att IT-kunskaperna skiljer sig betydligt mellan de olika individerna. Rapporten skall skrivas på dator och den skall dessutom skrivas i en mall. I mallen skall sedan bilder och tabeller läggas till. Studenterna ges ingen IT-undervisning och för en del studenter blir den språkliga rapportskrivningen lidande av att de inte besitter tillräckliga kunskaper i IT. För många i Sverige är det självklart att man behärskar de vanligaste mjukvarorna såsom Word, Excel och PowerPoint och man tänker inte på att många invandrare, även de med akademisk examen, inte är bekanta med datorer.

Studenterna skall inför sitt skrivande söka efter relevant information inför sin rapportskrivning på temat hållbar utveckling. Studenterna har på lägre nivå examinerats på ämnet hållbar utveckling och skall vara familjära med vad detta begrepp betyder. Studenterna har dock inte fått någon föreläsning eller workshop om hur man söker efter litteratur eller

träning i källkritik. Rapporten kan fälla de studenter som kommer från akademiska kulturer där inte källkritik är ett aktivt inslag under deras utbildning.

Kane (1992) och Messick (1989) har båda med relevansens betydelse för testets giltighet och Bachman och Palmer (2010) anser även att testkonstruktörerna och examinatorerna måste besitta kunskaper inom denna domän för att kunna belägga validiteten. Domänen som testas genom Invandrarakademins slutrapport är högskole-studiedomänen, vilken examinatorerna har kunskap inom. Studentens arbete skall innefatta en koppling till studenternas egen profession liksom till hållbar utveckling. Examinatorerna är båda svenska som andraspråks- och samhällskunskapslärare och av naturliga orsaker är det omöjligt för dem att besitta specialistkunskaper inom alla studenternas olika professioner såsom elingenjörer, logistiker och så vidare. Bachman och Palmer menar att tolkningarna av testresultaten blir meningslösa (2010) ifall testkonstruktören inte har tillräckliga kunskaper inom den testade domänen. I detta fall får Bachman och Palmer inte medhåll från mig eftersom jag anser att i detta fall skall rapporten bedömas utifrån att den är på gymnasial nivå eller inte och att fokus är på det svenska språket.

Språket i rapporten bedöms med hjälp av språkskalan GERS som utformades för att vara en självskattning för studenter av främmande språk under 70-talet. Språkkunskaperna för de olika europeiska invånarna skulle kunna vara jämförbara. GERS var inte tänkt att användas i andraspråksundervisning och dess bedömning och det finns svårigheter att använda en matris som är gjord för att bedöma ett annat ämne, anser jag, något jag inte får medhåll i från examinatorerna. En annan svårighet, anser jag, är att en stor del av bedömningen görs genom erfarenhet, vilket kan medföra att bedömningen blir subjektiv. Baker (2012) varnar för att examinatoren kan påverkas av konsekvenserna av bedömningen i sin bedömning liksom att det finns socio-kognitiva skillnader bedömare emellan.

Är det produkten eller processen som bedöms? Produkten är det som skall bedömas, men hur påverkar det att examinatorerna också är handledare och följer studenterna genom hela skrivprocessen från synopsis till färdig rapport. Dessutom så står det i innehållet för kursen att processen är i fokus. Examinatorerna anser själva att slutrapporten är en formativ bedömning, medan jag anser att slutrapporten bedöms genom att en summativ bedömning görs genom att det är slutprodukten som bedöms vara godkänd eller inte och inte processen. Examinatorerna följer studenterna genom hela processen, men det är endast slutprodukten som bedöms av de båda examinatorerna tillsammans.

Examinatorerna ger uttryck för att de känner tryck från studenterna att de skall ges godkänt på slutuppgiften eftersom flertalet studenters fortsatta akademiska karriär i Sverige försenas vid icke godkänt. Samtidigt uttrycker informanterna att i vissa yrkesgrupper är inte ett professionellt utfört förvärvsarbete avhängigt av språket. Konsekvenserna kan inte helt bortses ifrån när det gäller bedömningens tillförlitlighet och konsekvenserna i ett mikro- och ett makroperspektiv kommer alltid att finnas i denna studentgrupp oavsett konstrukt.

Pressen som lärarna känner och ger uttryck för gör att jag anser att det bör diskuteras bland lärarna ifall inte den sista examinationen bör genomföras av utomstående för att höja rättsäkerheten. Det går inte bortse ifrån subjektiviteten när examinatorerna även är handledare, undervisande lärare och testkonstruktörer. Institutionen för svenska och flerspråkighet vid Stockholms Universitet erbjuder ett väl utprovat behörighetsgivande test, TISUS, som är rättssäkert och gjort för denna målgrupp. Informanterna anser att TISUS är för brett för målgruppen, men de får inte medhåll ifrån mig då jag anser att studenternas kunskaper i svenska språket måste vara breda för ett Sv2/SVA2-betyg samt för att klara studier på svenska vid universitet eller högskola.

Av naturliga orsaker kan inte en individs hela språkliga förmåga bedömas utifrån en examination, vilket påpekas av Lundahl (2011) utan det är bara fragment som examinatorn/erna skall dra slutsatser ifrån. Detta medför att konstrukten måste vara genomtänkt ur flera perspektiv, både fysiska och psykiska. Jag anser att Invandrarakademins slutrapport har en relativt hög giltighet utifrån ovanstående diskuterade variabler.

6.2 Metoddiskussion

I denna rapport har en kvalitativ metod används vilken har genomförts i form av en valideringsstudie (materialgenomgång) samt två stycken intervjuer.

När man gör personliga intervjuer skall man vara medveten om att de intervjuade personerna kanske vill vara till lags och svarar det som de tror att man som intervjuare vill höra. Informanterna kan svara som de anser att de borde svara i den roll som de besitter som kursansvariga och examinatorer för kursen. Av egen erfarenhet vet jag att man inte vill komplicera saker och ting utan gärna vill svara enkelt och kortfattat på frågorna, ifall det inte är enkelt att förklara kan man ta en genväg och modulera sanningen. Det kan vara så att informanterna muntligen framförde vad de förväntade sig att jag ville ha för svar och de förenklade svaren för att de kanske ville att intervjun snabbt skulle bli

avklarad. Säljö (2005) menar att allt vi säger är tolkningar av omvärlden utifrån hur vi har blivit medierade från barnsben, vilket medför att analysen av intervjuerna inte bara är tolkningar av informanternas svar på frågorna utan också är min tolkning utifrån min mediering av informanternas svar på mina frågor.

Materialgenomgången består av allt befintligt material, både skriftligt och muntligt, som finns till förfogande för studenterna på lärplattformen. Men det kan också finnas utdelat material som inte laddats upp på lärplattformen som jag inte har fått ta del av. Det kan också finnas input till studenterna som har framförts muntligen då kameran varit avstängd till exempel i början av föreläsningen eller i slutet av föreläsningen. Materialet som har gått igenom kan förändras och kommer förmodligen bytas ut till nästkommande terminer, vilket påverkar slutsatsens generaliserbarhet till kommande slutrapporters giltighet gällande bedömning.

6.3 Slutsats

Grundlagen som är gällande för alla statligt anställda lärare vid ett universitet eller en högskola är att bedömningen som görs vid examinationer skall vara saklig och opartisk samt att allas likhet inför lagen skall beaktas. Genom att använda olika sorters bedömningsanalysinstrument och bedömningsmodeller kunde Invandrarakademins slutrapport analyseras utifrån olika perspektiv och på så sätt ses ur ett helhetsperspektiv.

Mina slutsatser är att slutuppgiften som Invandrarakademins studenter gör för att bli godkända på Sv2-nivå har en relativt hög giltighet och även tillförlitlighet eftersom dess relevans är hög utifrån att det är en högskoleförberedande kurs. Det som degraderar giltigheten är att GERS används som bedömningsmatris och inte någon specifik bedömningsmatris är skriven för denna examination. GERS är utarbetad för att användas i ämnet svenska som främmande språk och inte i svenska som andraspråk, vilket är ämnet som bedöms i Invandrarakademins. Dessutom är GERS en referensram för formativ bedömning, medan Invandrarakademins slutuppgift bedöms genom en summativ bedömning i och med att det endast är slutprodukten som bedöms genom ett godkänt eller icke godkänt.

Testets tillförlitlighet vägs upp av att det är två examinatorer som bedömer examinationsuppgiften tillsammans och att alla studenterna genomför samma uppgift. Språk är ingen isolerad ö, utan används i en social kontext och de sociala värden som examineras implicit och

explicit är relevanta i den högskoleförberedande kursen Invandrarakademin. Konsekvenserna i ett mikro- och makroperspektiv kan inte bortses ifrån när det gäller testets reliabilitet, men påverkar inte testets validitet enligt min åsikt.

Litteratur

- Bachman, L. F. & Palmer, A. S. 2010. *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Baker, B. A. 2012. Individual Differences in Rater Decision-Making Style: An Exploratory Mixed-Methods Study. *Language Assessment Quarterly*, 9:3, 225-248.
- Björkdahl Ordell, S. 2007. Vad är det som styr vilka etiska regler som finns? I: Dimenäs, J (red.), 2007. *Lära till lärare*. Stockholm: Liber, s 21-28.
- Common European framework of reference for languages (CEFR): learning, teaching, assessment 2001. Cambridge: Cambridge University Press.
- Eklund, G. 2011. *Kvalitativt inriktad metodologi och metod*. [pdf]. www.vasa.abo.fi/users/.../Met%20&%20Met%20Kompend.doc. Hämtad: 2013-03-16.
- Fulcher, G & Davidson, F. 2007. *Language testing and assessment*. London & New York: Routledge.
- Huang, S. 2012. Pushing Learners to Work Through Tests and Marks: Motivating or Demotivating? A case in a Taiwanese University, I: *Language Assessment Quarterly*, 9:1, 60-77.
- Högskolan i Borås 2012. *Svenska som andraspråk för utländska akademiker för ett hållbart samhälls- och arbetsliv; grundkurs 31-60 hp*. <http://www.hb.se/Vill-studera/Program-och-kurser/Kurser-HT-2013/Svenska-som-andrasprak-for-utlandska-akademiker-for-ett-hallbart-samhalls--och-arbetsliv-grundkurs-31-60-hp/#syllabus>. Hämtad: 2013-06-12.
- Kane, M. 1992. An argument-based approach to validity. I: *Psychological Bulletin*, 112:3, 527-535.
- Lundahl, C. 2011. *Bedömning för lärande*. Stockholm. Nordstedts.
- McNamara, T. F. & Roever, C. 2006. *Language testing: the social dimension*. Oxford: Blackwell.
- Messick, S. 1989. Validity. I: Linn, Robert L. (red.), *Educational measurement*. 3. ed. New York: American Council on Education.
- Regeringsformen, kap 1, §9. *Kungörelse (1974:152) om beslutad ny regeringsform*. [elektronisk] <https://lagen.nu/1974:152>. Hämtad: 2013-05-11.
- Skolverket 2009. *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning*. Stockholm: Skolverket

- Sundberg, G. 2007. Assessing and testing second language proficiency in the Nordic countries. I: Carlsen, Cecilie & Moe, Eli (red.), *A Human Touch to Language Testing*. Novus Press. S. 87–96.
- Sundberg, G. 2010. Språkbeskrivningar i ett andraspråksperspektiv – språket sett som regel eller som resurs? I: Olofsson, Mikael. (red.), *Symposium 2009. Genrer och funktionellt språk i teori och praktik*. Stockholm: Stockholms universitets förlag. S. 292–302.
- Säljö, R. 2005. *Lärande & kulturella redskap*. Falun: Nordstedts akademiska förlag.
- Van Avermaet, P. 2009. Fortress Europe? Language policy regimes for immigration and citizenship. I: Hogan-Brun, G, Mar-Molinero, C, Stevenson, P (red.), *Discourses on Language and Integration*. Amsterdam/Philadelphia: John Benjamins B.V.S.
- Vetenskapsrådet 2005. Vad är god forskningssed? Synpunkter, riktlinjer och exempel. [pdf]. <http://www.vr.se/download/18.6b2f98a910b3e260ae28000334>. Hämtad: 2013-05-12.
- William, D. 2013. *Att följa lärande: formativ bedömning i praktiken*. 1. uppl. Lund: Studentlitteratur

Bilaga 1 – intervjufrågor

1. Hur definierar du rättssäker bedömning?
2. Motivera val av examination?
3. Hur är relationen mellan studenten och lärare? Studenter och bibliotikarier?
4. Vilken språksyn har du haft i testkonstruktionen?
5. Vad är konstrukten?
6. Vad är den förväntade inputen?
7. Vad är den förväntade outputen?
8. Förklara testets relevans?
9. Förklara resultatens generaliserbarhet?
10. Hur sker bedömningen?
11. Vad bedöms?
12. Vilka konsekvenser ser du av ett godkänt eller icke godkänt?
13. Varför används inte TISUS?