

**Att bedöma skriftliga färdigheter i svenska med hjälp av en  
nivåskala:**

**hur skillnader mellan bedömare uppstår**

Kandidatavhandling

Miia Juutilainen

Jyväskylä universitet  
Institutionen för språk  
Svenska språket

28.8.2011

## TIIVISTELMÄ

### JYVÄSKYLÄN YLIOPISTO

Humanistinen tiedekunta	Kielten laitos
Tekijä: Juutilainen, Miia	
Otsake: Att bedöma skriftliga färdigheter i svenska med hjälp av en nivåskala: hur skillnader mellan bedömare uppstår	
Aine: ruotsin kieli	kandidaatintutkielma
Vuosi 2011	Sivumäärä 40
<p>Tutkielman tarkoituksena on selvittää, kuinka neljä kokenutta kielitaidon arvioijaa arvioi suomenkielisen yläasteopiskelijan kirjoittaman ruotsinkielisen kirjoitelman käyttäen apunaan Opetushallituksen kielitaidon tasojen kuvausasteikkoa. Tavoitteena on selvittää, missä määrin arvioijat tukeutuvat taitotasosteikkoon arvioidessaan kirjoitelmaa, millä asteikkoon kuuluvilla tai kuulumattomilla kriteereillä he useimmin perustelevat valintansa sekä miksi arvioijat arvioivat kirjoitelman eri tavoin.</p> <p>Analysoitavaksi kirjoitelmaksi valittiin sellainen, jolle arvioijat määräsivät eri tasoja: kaksi päätyi A2.1:een, yksi A1.3:een ja yksi A2.2:een. Tutkimusmetodinä käytettiin puolistrukturoitua teemahaastattelua sekä haastatteluaineiston kvalitatiivista analyysia.</p> <p>Tulokseksi saatiin, että kaikki arvioijat perustelivat tasovalintansa sekä sellaisilla kriteereillä, jotka kuuluivat tasokuvauksiin että sellaisilla perusteilla, jotka eivät suoranaisesti liittyneet taitotasosteikkoon. Suuria eroja siinä, missä määrin arvioijat tukeutuivat tai eivät tukeutuneet taitotasosteikkoon, ei ollut. Eroavaisuuksia syntyi lähinnä siinä, kuinka ratkaisevia asteikon ulkopuoliset kriteerit tietyn taitotason valinnalle olivat. Arvioijat perustivat tasovalintansa useimmin aikamuoto- ja verbintäivutusseikkoihin, ymmärrettävyyteen, virheisiin, sanastollisiin tekijöihin, alisteisten ja rinnasteisten lausetyyppien hallitsemiseen sekä muiden kielten siirtovaikutukseen. Huomattavaa on, etteivät sisällölliset tai kieliopilliset seikat hallinneet arvioijien perusteluja. Yleisimmin asteikon ulkopuolisista kriteereistä viitattiin siirtovaikutukseen, vaihtelevien lauserakenteiden käyttämiseen sekä ymmärrettävyyteen, joista jälkimmäiseen kaikki neljä arvioijaa tukeutuivat.</p> <p>Eroavaisuudet taitotason määrittämisessä selittyvät sillä, että arvioijat tulkitsivat taitotasokuvauksia eri tavoin. Arvioijat myös painottivat vaihtelevassa määrin joko kirjoitelman vahvuuksia tai heikkouksia siten, että vahvuuksia painottavat arvioijat määräsivät kirjoitelmalle paremman taitotason kuin heikkouksia painottava. Mahdollista myös on, että yksi arvioijista jätti huomiotta sellaisen myönnytyksen taitotasokuvauksissa, joka oli ratkaiseva kolmen muun arvioijan ratkaisulle.</p> <p>Tutkielman pääasiallinen päätelmä on, että vaikka arvioijat arvioivat kirjoitelman vaihtelevin taitotasoin, perustelivat he valintansa suhteellisen yhdenmukaisesti. Pienet erot tasokuvauksien tulkinnassa sekä vahvuuksien tai heikkouksien painottamisessa aiheuttivat kuitenkin sen, etteivät arvioijat olleet yhdenmukaisia annetun tason suhteen.</p>	
Avainsanat: bedömning, elevbedömning, evaluering, skrivfärdighet, språkprov, uppsatsskrivning, tillämpad språkvetenskap, utvärdering	
Kirjasto/Säilytyspaikka:	
Muita tietoja:	

## Innehåll

1	INLEDNING.....	4
2	TEORIBAKGRUND .....	5
2.1	Språkbedömning och språktest.....	6
2.2	Nivåskala för språkkunskaper som stöd för bedömning .....	7
2.3	Bedömaren.....	11
2.4	Tidigare forskning .....	13
3	MATERIAL OCH METOD .....	16
3.1	Källan till materialet .....	17
3.2	Nivåskala för språkkunskaper .....	18
3.3	Informanterna .....	19
3.4	Det praktiska materialinsamlandet .....	20
3.5	Intervjuerna.....	21
3.6	Analysen .....	22
4	RESULTAT .....	23
4.1	Hur uppsatsen bedömdes .....	24
4.2	Vilka egenskaper som diskuterades.....	27
4.3	Hur bedömarna använde nivåskala för språkkunskaper .....	29
4.4	Vilka likheter och skillnader det uppstod .....	31
5	DISKUSSION.....	35
5.1	Forskningsfråga 1 .....	35
5.2	Forskningsfråga 2 .....	36
5.3	Forskningsfråga 3 .....	37
5.4	Forskningsfråga 4 .....	38
5.5	Om tillförlitligheten.....	40
6	AVSLUTNING.....	41
	LITTERATUR .....	44
	BILAGOR.....	46

## 1 INLEDNING

Språkbedömning har en ofrånkomlig roll vid inläring av ett språk. Enligt Tarnanen (2003) syftar man genom att sätta betyg till att ge återkoppling om inlärares egenskaper både till inlärares själv och till grupper som står bakom bedömningen. Vidare konstaterar Tarnanen, Huhta och Pohjala (2007) att det är just betyget och de tilldelade nivåerna som ofta ligger till grund för evaluering av undervisning i klassrummet, i skolan eller i hela landet. Betygen och nivåerna spelar även en stor roll vid elevtagning till högskolor och andra läroanstalter. Med tanke på bedömningens stora inflytande på inlärares och på olika verksamheter ställer forskarna stora krav på bedömningens kvalitet: bedömningsresultaten förutsätts vara pålitliga, trovärdiga och etiska samt utvisa inlärares verkliga kunskaper.

I de flesta undersökningar om språkinläring står eleven och inlärningsresultaten i fokus. Mindre uppmärksamhet har riktats mot bedömaren. Enligt Tarnanen, Huhta och Pohjala (2007) kan den dåliga kännedomen om ämnet bero på att den vetenskapsgren där man forskar om språkbedömning är relativt ung och erfarna experter inom fältet är få. Forskarna påpekar att följden är att till och med lärarna själva har dåliga kunskaper om hur man bedömer språkliga färdigheter. Tarnanen, Huhta och Pohjala lyfter fram att flera undersökningar har visat att betygsättning i skolor inte är enhetlig och att elever med samma betyg kan ha betydande olikheter i språkkunskaper.

Utöver det faktum att studier om språkbedömning är relativt få i förhållande till undersökningar om andra aspekter inom språkinläring har de forskares intresse som studerat bedömningsprocessen fördelat sig mellan olika delområden i språkfärdighet, såsom muntlig eller skriftlig kompetens, mellan olika språk eller mellan första- och andraspråk. Studier om hur man bedömer svenskspråkiga texter skrivna av finskspråkiga elever är ett ämne som forskare sällan gripit sig an.

I denna avhandling kommer jag att utreda hur fyra bedömare bedömer en uppsats på svenska skriven av en finsk högstadiellev med hjälp av läroplanens nivåskala för språkkunskaper. Mitt största intresse inriktar sig inte på tilldelade nivåer och i vilken grad bedömarna är systematiska och ense om nivån, utan jag är intresserad av om det finns skillnader i olika bedömares sätt att bedöma och använda nivåskalan och om detta påverkar

bedömningsresultaten. Studiens syfte är således att förklara varför bedömarna inte är ense om nivåerna trots att de använder en gemensam nivåskala för språkkunskaper. Dessutom skall jag redogöra för vilka egenskaper i uppsatsen bedömarna fäster avseende vid och om dessa egenskaper tillhör nivåskalan. Till sist skall jag försöka utreda om skillnader i bedömningarna kan förklaras med bedömarnas skilda sätt att beakta olika element i uppsatsen eller med deras olika sätt att använda nivåskalan. Därmed skall jag formulera mina forskningsfrågor på följande sätt:

1. I vilken mån använder bedömarna nivåskala för språkkunskaper (Läroplansgrunderna 2004) när de bedömer uppsatsen?
2. Vilka kriterier använder bedömarna mest i sina bedömningar?
3. Finns det kriterier som bedömarna använder i bedömningarna men som inte tillhör nivåskalan?
4. Har bedömarnas sätt att bedöma eller använda nivåskalan ett samband med de angivna nivåerna?

För att kunna besvara dessa frågor kommer jag att analysera fyra intervjuer där bedömarna själva berättar hur de kom fram till vissa angivna nivåer när de bedömde uppsatsen. För analysmetoden, materialinsamlandet, bedömarna, uppsatsen och den använda nivåskalan redogörs närmare i kapitel 3. Med hjälp av iakttagelser som framkommit vid analysen av intervjuerna kommer jag i kapitel 4 att framlägga resultaten av analysen. Upptäckterna och deras samband med de angivna nivåerna diskuterar jag i kapitel 5, där jag skall ge en återblick till de ovannämnda forskningsfrågorna. För att kunna definiera de viktigaste termerna och presentera de väsentligaste teorierna för denna avhandling samt utreda några tidigare forskningsresultat i ämnet skall jag emellertid först ge en översikt över den teoretiska bakgrunden i det följande kapitlet.

## **2 TEORIBAKGRUND**

I detta kapitel skall jag presentera de viktigaste termerna i och teorierna för denna studie. Jag inleder med att i avsnitt 2.1 definiera begreppen språkbedömning och direkt språktest samt med att presentera de mest relevanta egenskaper som kännetecknar ett bra språktest. I avsnitt

2.2 redogör jag för hur och varför nivåskala för språkkunskaper används vid holistisk bedömning samt lyfter fram några problem vid användning av holistiska nivåskalor. Hur bedömaren påverkar bedömningsresultatet diskuteras i avsnitt 2.3 i samband med presentation av några typer av de vanligaste bedömarberoende mätningssfelen. Till slut redogör jag kort för några tidigare forskningsresultat i avsnitt 2.4.

## 2.1 Språkbedömning och språktest

För att kunna granska det som händer vid bedömning måste man först ha klart för sig vad det innebär att bedöma språkliga färdigheter. I denna studie avser jag med begreppet *språkbedömning* hur man med direkta skrivtest bedömer inlärares *färdighetsnivå* (proficiency; begrepp introducerat av Huhta och Takala 1999) till skillnad från den typ av bedömning som anknyter till inlärares kunskap om språket. Enligt Huhta och Takala (1999:189–190, min översättning) gäller det att utforska ”inlärares aktuella absoluta kunskapsnivå (osaamisen taso) och därmed *tillräckligheten av färdigheter* (taitojen riittävyys) för ett visst ändamål” när man bedömer testtagarens färdighetsnivå. Enligt Weigle (2002:46) talar man då om *prestationsbedömning* (performance assessment). Med begreppet prestationsbedömning avser Weigle sådana bedömningsaktiviteter som antingen innebär observationer av beteende i verkligheten eller iakttagelser av sådana aktiviteter som simulerar verkligheten. I min studie antar jag att språkbedömning representerar den senare typen när man förmodar att man av en text som testtagaren har producerat kan dra några slutsatser om dennes verkliga språkförmåga. Termen *verklig språkförmåga* använder jag i denna undersökning för att åsyfta kunskaper som avgör i vilken grad språkinlärares språkligt klarar sig i sitt vardagliga liv, i sina studier och i sitt arbete på språket ifråga (jfr Council of Europe 2009).

Färdighetsnivå i ett främmande språk kan i princip testas och bedömas under samma villkor som bedömning av annat utförande när testet innehåller det egentliga skrivandet i motsats till bl.a. flervalsuppgifter. I Council of Europe (2009:184–185) kallas ett språktest där testtagaren verkligen producerar text *ett direkt test*. Enligt denna definition fungerar texten som testtagaren skrivit som prov på testtagarens skriftliga kompetens, dvs. hur väl språkinlärares kan producera språk. Därmed är det primära syftet med ett direkt språktest enligt den europeiska referensramen att kunna dra generaliserade slutsatser om färdigheter i språket på

basis av testprestationen. Först därefter skall man kunna fatta beslut som baserar sig på dessa slutsatser och som enligt min tolkning fungerar som grund för tilldelade nivåer.

Ett språktest är användbart om det uppfyller vissa krav på rättvishet och pålitlighet. Bachman och Palmer (refererad i Weigle 2002) lyfter fram egenskaper som reliabilitet, validitet, autenticitet, interaktivitet, testeffekt och genomförbarhet, varav jag anser reliabilitet och validitet vara de mest relevanta för denna undersökning. *Reliabilitet* i ett skrivprov innebär att om man testar inlärarna två gånger får man samma resultat båda gångerna och att olika bedömare tilldelar samma nivåer. Brister på reliabilitet kan uppstå när en bedömare bedömer texter av samma kvalitet inkonsekvent eller hans/hennes bedömningar av en och samma text osystematiskt varierar från den ena gången till den andra. En annan anledning till opålitlighet i ett skrivtest är otillräcklig interbedömarreliabilitet som vållas av oförenhetlighet mellan bedömarna. (Weigle 2002:48-49; Huhta och Takala 1999:211-212; Lindroos 2010:33).

Med begreppet *validitet* åsyftas till i vilken grad testet mäter det som den har för avsikt att mäta (Weigle 2002:49), eller hur väl motiverade tolkningar av testresultaten är (Huhta och Takala 1999:213). Om man t.ex. vill mäta inlärarens skriftliga färdigheter i svenska måste man först ha klart för sig vad som avses med skriftliga färdigheter. Betyder en hög skriftlig kompetens till exempel grammatisk felfrihet eller ett stort ordförråd i svenska språket, eller något helt annat? Sedan måste man noggrant överväga om testet verkligen testar dessa kunskaper. Om man t.ex. antar att en bra skriftlig färdighet i svenska åtminstone innebär förmågan att binda samman enstaka ord till en sammanhängande, lättläst text, då skulle en flervalsuppgift inte ge möjlighet att studera denna kunskap och vore således inget bra val för ett skrivprov. Dessutom, för att kunna skapa en hög grad av validitet i ett språktest är det viktigt att de gemensamma kriterierna är utvalda så att de stödjer den uppfattning om språket som anses vara relevant i språkinläring, och att bedömarna håller fast vid dem (Weigle 2002).

## **2.2 Nivåskala för språkkunskaper som stöd för bedömning**

Nivån tilldelad en text påverkas inte enbart av språkinlärarens personliga karaktärsdrag och skriftliga färdigheter utan också av ett flertal faktorer kring bedömaren, använda bedömningsstrategier och nivåskalor samt av externa omständigheter som medverkar i

beslutsfattandet (se t.ex. Weigle 2002). På grund av att syftet med denna studie är att utreda hur bedömarna stödjer sig på nivåskala för språkkunskaper bildar emellertid bedömaren och färdighetsskalan grunden för undersökningen.

Förmågan att skriva bra svenska är ett mycket abstrakt begrepp och kan därmed inte helt entydigt definieras. Därför kan man ha olika synsätt på vilka egenskaper som hör till en bra svensk text skriven av en finsk högstadiellev. För att kunna jämföra olika test med varandra och för att försäkra sig om att bedömaren baserar sina bedömningar på samma kriterier är det vanligt att använda en *nivåskala*. I denna studie använder jag begreppet nivåskala för att syfta på en stigande skala av nivåer som är utrustad med beskrivningar som bestämmer vilka språkliga färdigheter inläraren borde äga för att kunna bli tilldelad en viss nivå. Dessa beskrivningar kallas i min undersökning *nivåbeskrivningar* eller *kriterier* (jfr dock Tarnanen 2003:150). Nivåbeskrivningar kan innehålla både negativt laddade framställningar som betonar fel och testtagarens brister (t.ex. ”på grund av en bristfällig behärskning av vokabulär uppstår det missförstånd”) och mer positiva beskrivningar av en exemplarisk prestation på den ifrågavarande nivån (t.ex. ”kan producera korta vardagliga uttryck”). Den ena ändan av skalan beskriver en hög skriftlig kompetens medan den andra beskriver kunskaper som förväntas av en nybörjare. Däremellan återfinns en rad stigande nivåer vars krav på språkfärdighet varierar beroende på hur högt inläraren befinner sig i nivåskalan och som motsvaras av var sin beskrivning av kvalitet på språkbehärskning. Bedömningen kan på så sätt genomföras genom att jämföra testprestationen med nivåbeskrivningar i nivåskalan och på basis av detta placera testtagaren på den nivå som bäst motsvarar kvaliteten av prestationen. Ju mer texten innehåller egenskaper av bra skrivande desto högre nivå blir den tilldelad. (Se t.ex. Council of Europe 2009; Huhta och Takala 1999; Tarnanen 2002).

I nivåbaserad bedömning är det inte ändamålsenligt att jämföra inlärarens prestationer med kunskaper hos andra testtagare, utan bedömningen av en enskild individ skall vara oberoende av hur väl testgruppen i sin helhet lyckats i testet eller av rangordningen mellan testtagarna (Council of Europe 2009). I bedömning som är bunden till nivåbeskrivningar är det således fråga om att mäta inlärarens språkliga kompetens, något som inte kan genomföras utan att så exakt som möjligt precisera vad språkfärdighet i just det förevarande testet innebär (Huhta och Takala 1999; Tarnanen 2003). Därför har nivåbeskrivningar en klar anknytning till testorganisatorns uppfattningar om språket och vad som är det viktigaste att lära sig i

främmande språk. Med tanke på detta står det klart att nivåbeskrivningarna skall vara utformade i enhetlighet med testets målsättning.

Nivåbeskrivningarna i denna studie (Läroplansgrunderna 2004) bildar en holistisk skala, vilket innebär att varje färdighetsnivå motsvaras av en enda nivåbeskrivning. Språkkunskaper är således inte uppdelade i systematiska delar som bedöms skilt från varandra så som meningen är vid användning av analytiska skalor (Tarnanen 2003:150). Enligt Tarnanen (2002; 2003) kan holistiska nivåbeskrivningar ofta innehålla olika kriterier på t.ex. rättskrivning, satsernas längd eller vokabulär men trots det förväntas bedömaren inte poängsätta dessa delområden vid holistisk bedömning. Bedömaren ger däremot ett sammansatt helhetsbetonat betyg för texten genom att intuitivt evaluera olika delar av prestationen. I det följande skall jag ge ett exempel på en holistisk nivåbeskrivning som ingår i nivåskalan utformad av Utbildningsstyrelsen (7.8.2011). Nivåkriterierna är avsedda för att beskriva egenskaper hos en inlärare som på basis av hans/hennes skriftliga prestationer befinner sig på den lägsta nivån av skalan, nivån A1.1:

- Kan delge sina omedelbara behov i mycket korta ordalag.
- Kan skriva språkets bokstäver och siffror med bokstäver, anteckna sina personuppgifter och skriva vissa bekanta ord och fraser.
- Behärskar ett antal enskilda ord och uttryck.
- Är inte förmögen att uttrycka sig fritt men kan skriva vissa ord och uttryck rätt.

Tarnanen (2002:85–87) konstaterar att det även vid holistisk bedömning är nästan omöjligt att hålla de analytiska och holistiska bedömningsmetoderna isär. Tarnanen anser att detta är en följdverkan av att även de holistiska nivåskalorna åtminstone delvis oundvikligen brukar bli analytiska till sin natur. Kriterier som i viss grad kan anses präglas av analytiska drag och som ofta framträder i holistiska nivåbeskrivningar kan angå egenskaper såsom textens längd, ordförråd eller förståelighet. Ett exempel på en sådan holistisk nivåbeskrivning som enligt Tarnanen (2002:87, min översättning) åtskiljer dessa drag är följande: ”kan producera en kort bit av text som består av lösryckta ord och som är svår att förstå”. Det är då längden, ordförrådet och förståeligheten som bedömaren enligt Tarnanen mest uppmärksammar även om han/hon borde bedöma holistiskt. Således föreslår Tarnanen att en holistisk bedömningsprocess i grunden är en sammansättning av holistiska och analytiska egenskaper. Med detta avser Tarnanen att bedömaren ger texten en holistisk nivåbedömning men kommer fram till detta resultat genom att analytiskt värdesätta olika aspekter i texten.

Nivåbeskrivningarnas funktion i holistisk bedömning är enligt Tarnanen (2002; 2003) att tjänstgöra som allmän instruktion för att hjälpa till med att inrikta bedömarens uppmärksamhet till relevanta egenskaper i texten. Dessutom skall nivåskalan enligt Tarnanen garantera att bedömaren betonar och lägger märke till olika språkliga delområden på samma sätt utan att koncentrera sig på bara några enstaka kriterier. Utan nivåskalan skulle de tilldelade nivåerna med andra ord bli mer beroende av enskilda bedömares åsikter och uppfattningar och testet skulle komma att sakna utgångspunkter för en enhetlig bedömning. Genom att använda en nivåskala är det således möjligt att förminska inverkan av bedömarens subjektiva bedömningssätt på testresultaten och därmed styra bedömningen från intrycksbaserad till en mer kontrollerad verksamhet (Council of Europe 2009; jfr dock Tarnanen 2002:84–85). Genom kontrolleringen når man en högre grad av enighet och tillförlitlighet i testresultaten.

Erfarenheter i praktiken har emellertid visat att det inte är helt oproblematiskt att använda nivåskalor i synnerhet vid holistisk bedömning. På grund av den allmänna naturen av en holistisk nivåskala för språkkunskaper lämnar skalan enligt Tarnanen (2002:85–87) utrymme för olika tolkningar. Problem med oenighet och otillförlitlighet kan enligt Tarnanen (2002:76–77) uppstå i synnerhet om nivåbeskrivningarna inte är väl formulerade. Enligt Tarnanen är nivåskalorna ibland sammanställda så att de skildrar färdigheterna bara på ett allmänt, flertydigt eller oklart sätt eller så att färdigheten beskriven i skalorna motsägs av färdigheten krävd i själva uppgiften. I båda fallen ökar enligt Tarnanen risken att bedömaren baserar sina bedömningar på sina egna tolkningar i stället för på nivåkriterierna.

Problemet med flertydiga nivåbeskrivningar kan förstås när man betraktar den komplexa egenarten av holistiska nivåskalor. Tarnanen (2002; 2003) anser att alla kriterier som påverkar beslutsfattandet i bedömningen inte kan komprimeras i nivåbeskrivningar. Detta belyser hon med ett exempel på hur skriftlig kompetens inte ur en holistisk synvinkel kan anses bestå av en summa av t.ex. vokabulär, språkliga strukturer, förståelighet och koherens. Holistisk syn på bedömning grundar sig däremot enligt forskaren på principen att språklig färdighet är någonting mer än summan av dess beståndsdelar.

Trots de ovan diskuterade särdragen vid holistisk bedömning betonar Tarnanen (2002; 2003) att bedömaren inte hur som helst kan tilldela nivåer: det holistiska intrycket av uppsatsen skall inte bli subjektivt utan basera sig på nivåbeskrivningar i nivåskalan.

### **2.3 Bedömaren**

Trots att bedömaren har en nivåskala för språkkunskaper som stöd för sin bedömning är det slutligen bedömarens övervägande som bildar grunden för de tilldelade nivåerna. Därmed är det bedömaren som styr bedömningsprocessen antingen medvetet eller omedvetet och vars uppfattningar om skriftliga färdigheter, vars fysiska och mentala tillstånd och vars erfarenheter samverkar i processen. Det föreligger flera externa och fysiska faktorer hos bedömaren, såsom trötthet, buller, erfarenhet, ovana och utbildning eller brist på utbildning, som kan påverka bedömningsresultatet (Tarnanen 2002; 2003). Listan av samverkande komponenter under en bedömningsprocess kan kompletteras av t.ex. antalet och ordningen av texter som skall bedömas samt känslan av tidsbrist (Weigle 2002) och denna lista verkar på basis av undersökningar fortgå i oändlighet. I föreliggande undersökning skall jag emellertid inte gå närmare in på detaljer av ytliga faktorer som påverkar bedömningsresultatet utan jag skall sätta mig in i sådana olikheter i bedömares sätt att bedöma som beror på bedömares individuella val i en bedömningsituation och dennes individuella bedömningsstil.

Tarnanen (2002; 2003) framhäver att en bedömare är en individ vars bedömning alltid kännetecknas av en viss grad av subjektivitet. Sådana mentala faktorer som samverkar i bedömningen är enligt Tarnanen bl.a. uppfattningar och kunskap om språket och världen, värdesättningar, erfarenheter samt föreställningar och förväntningar kring texten som skall bedömas. Vidare anser Tarnanen att det alltid finns faktorer i bedömningen som bedömaren inte är medveten om eller inte kan påverka. Enligt Tarnanen medverkar dessa individuella faktorer i beslutsfattandet i viss mån även om bedömningen baserar sig på en gemensam nivåskala för språkkunskaper.

För att kunna garantera att testet uppfyller de strikta kraven på etiskhet och för att kunna bemöta testtagarna med rättvisa borde bedömningen vara så objektiv som möjligt. Därför borde man enligt Council of Europe (2009) försöka minimera sådana individuella värdesättningar hos en bedömare som leder till subjektiva beslut angående testtagarens

prestationer. Enligt Weigle (2002) ligger problemet med bedömning av direkta skrivtest emellertid i att några bedömare oftare lägger märke till olika språkliga drag än andra bedömare. Observationer om inlärares språkbehärskning kan enligt Weigle vara ytliga, ensidiga eller slumpmässiga och därmed kan slutsatser om inlärares språkliga färdigheter bli felaktiga, vilket i sin tur kan leda till bristande följdriktighet och tillförlitlighet i testresultaten.

När oenigheten eller inkonsekvensen av bedömningarna beror på bedömarens och inte på uppgiften, nivåskalan eller andra externa faktorer talar man enligt Tarnanen om *bedömarberoende mätningssfel* (arvioijalähtöinen mittausvirhe). Tarnanen (2002:68–71; 2003:156) presenterar flera olika typer av bedömarberoende mätningssfel varav de som mest har signifikans i denna studie är bedömarens *stränghet* och *mildhet* (ankaruus och lempeys) samt *haloeffekt* (sädekehävaikutus).

Kännetecknande för en sträng bedömare är enligt Tarnanen (2002:68) att han/hon mestadels brukar använda den nedre ändan av färdighetsskalan medan en mild bedömare brukar hålla sig till den övre ändan. En sträng bedömare använder således enligt Tarnanen färdighetsskalan bara upp till mitten eller något ovanför mitten, medan skalan för en mild bedömare praktiskt taget börjar först i mitten av färdighetsskalan eller lite nedanför. Tarnanen framhåller emellertid att bedömarna sällan är helt stränga eller helt milda i sina bedömningar utan ofta är det så att stränga och milda bedömningar endast vittnar om en smärre benägenhet att betona någondera ändan av skalan. En påfallande omtolkning av skalan för färdighetsnivåer skulle enligt Tarnanen betyda t.ex. att en låg nivå blir en medelnivå eller tvärtom. I denna studie antar jag emellertid att en sådan avsevärd omtolkning av skalan inte kommer att förekomma på grund av informanternas långa erfarenhet (se avsnitt 3.3).

Problem med haloeffekt uppstår i sin tur enligt Tarnanen (2002:68–69) när bedömarens inte lägger märke till alla egenskaper som förutsätts i nivåskalan utan bedömningen styrs av enstaka inslag i texten. Vidare konstaterar Tarnanen att haloeffekt också kan vara en följd av bedömarens subjektiva preferenser för texten eller den bild av testtagarens bakgrund som texten förmedlar. I sådana fall där bilden för bedömarens framstår i gynnsamt ljus brukar han/hon enligt Tarnanen tilldela texten en högre nivå än i sådana fall där bedömarens av någon anledning ställer sig negativ till texten och/eller testtagaren bakom texten. Tarnanen lyfter fram t.ex. handstil som exempel på en sådan individuell affektion som kan vilseleda bedömarens till att tilldela texten en mer subjektivt baserad nivå och som på så sätt kan

medföra fenomenet haloeffekt. I denna studie är emellertid påverkan av handstilen och bedömarens individuella förhållande till inläraren minimerad genom att material som är maskinskrivet används och att testtagarnas identitet inte framkommer. Dock är det möjligt att bedömarna ändå i viss mån baserar sina bedömningar på den bild av texttagaren som förmedlas genom texten.

För att göra bedömningen av flera bedömare överensstämmande med varandra måste man enligt Weigle (2002) säkerställa att bedömarna verkligen stödjer sig på nivåkriterier genom att använda skalan ändamålsenligt och konsekvent. Dessutom betonar Tarnanen (2002) att bedömaren trots bedömningens subjektiva natur kan eftersträva att bedöma följdriktigt med väl motiverade val och genom att reflektera över sitt eget agerande. Ju mer medveten bedömaren är om faktorer som påverkar bedömningen desto bättre kan han/hon enligt Tarnanen styra sina handlingar.

## **2.4 Tidigare forskning**

I ljuset av tidigare undersökningar tycks bedömning i olika kontexter vara förbluffande överensstämmande med avseende på egenskaper som olika bedömare lyfter fram och uppmärksammar i sina bedömningar (t.ex. Vaughan 1991; Weigle 2002). Detta bekräftas av Tarnanen (2002:126–131) som konstaterar att iakttagelserna och betoningarna hos olika bedömare vanligen är ytterst likartade oberoende av nivåskalan som ligger till grunden för bedömningar. Enligt Tarnanen (2002) är det grammatik och strukturer som bedömaren oftast tar upp i texter. Tarnanen upptäckte dessutom att bedömarna tenderade att göra grammatikbaserade val i synnerhet när texten låg vid gränsen mellan två nivåer eller när bedömarna annars hade svårigheter med att bestämma nivån. Då kunde testtagarens bristfälliga eller tillräckliga grammatiska felfrihet enligt Tarnanen utgöra den avgörande effekten.

Även om observationer av strukturer verkar gälla mindre grammatiska kategorier såsom tempusformer, syntax och ordböjning, vilket enligt Tarnanen (2002) lämnar utrymme för bedömarberoende betoningar mellan olika strukturer och deras delområden, tycks olika bedömare koncentrera sig på samma slags grammatiska element i sina bedömningar. Dessa betoningar är enligt Tarnanen emellertid inte alltid helt neutrala i förhållande till alla de

grammatiska element som kunde lyftas fram i en text. Detta motiverar Tarnanen genom att hänvisa till en undersökning av Vaughan (1991), där det framkom att bedömarna ofta baserade sina bedömningar på behärskning av en viss struktur på bekostnad av andra grammatiska komponenter.

Grammatikens dominans i bedömning har enligt Tarnanen (2002) kritiserats hårt och det är därför många bedömare som försöker undvika att vara för grammatiskt orienterade. Egenskaper som bedömare utöver grammatiken brukar evaluera i texter kan enligt Tarnanen (2002:130) vara förståelighet, läsbarhet, innehåll, disposition och kohesion. Av dessa verkar innehåll enligt min tolkning vinna mest uppmärksamhet och mest detaljerade kommentarer från bedömare i Tarnanens (2002) studie, där innehållet beskrevs med flera varierande slags yttringar. Enligt Tarnanen hänvisas till ett tilltalande innehåll med adjektiv som *bra*, *stark* och *livfull* (hyvä, vahva och elävä), vilka som framkommit i flera undersökningar (t.ex. Vaughan 1991), medan ett mindre lyckat innehåll beskrivs med ord som *dålig*, *otydlig*, *svag* och *tråkig* (huono, epäselkeä, heikko och tylsä). Som Tarnanen hävdar kan problemet med dessa yttringar anses vara deras ytlighet och subjektivitet.

En ytterligare kategori som bedömare ofta lägger märke till vid bedömning är texternas vokabulär. Tarnanen (2002:129–130) konstaterar att det vanligaste sättet att iakttä vokabulären är att granska omfånget av ordförrådet och evaluera om testtagaren har ett tillräckligt omfattande ordförråd för att kunna klara sig i testuppgiften. Andra aspekter kring vokabulären som Tarnanen nämner är variation, precision, idiomatiskhet samt lämplighet.

Charney (refererad i Vaughan 1991) poängterar att bedömning är en onaturlig verksamhet. Hon förklarar att bedömare på grund av det tillgjorda väsendet av bedömning fastnar i sådana element i texter som är lätta att hänvisa till men som tyvärr inte är betydelsefulla för riktiga skriftliga färdigheter. Detta anser Charney vara en orsak till att bedömare ofta är så enhetliga i sina bedömningar. Sådana ytliga egenskaper som bedömare grundar sina bedömningar på är enligt Charney handstil, ordval, textens längd och felskrivningar. Många studier verkar vederlägga Charneys påståenden. T.ex. Stewart och Grobe (refererade i Vaughan 1991) upptäckte att egenskaper som mest påverkade bedömningsresultat var längden på texterna samt avsaknaden av enkla mekaniska fel. Dessutom märkte Vaughan (1991) att bedömare inte bara tenderade att betona ytliga faktorer i sina bedömningar utan också att de jämförde texterna med varandra. Påfallande är att varken tonvikten på textens ytliga

egenskaper eller jämförandet av texterna hörde till den nivåskala eller de kriterier som användes i undersökningarna.

För att sammanfatta faktorer i texter som påverkar bedömningsresultat skall jag i det följande lista några egenskaper som bedömarna mest brukar hänvisa till i sina bedömningar. Listan grundar sig på en tabell som Tarnanen (2002:126–127) presenterar i sin avhandling och som i sin tur baserar sig på flera olika forskares undersökningar (Cumming; Vaughan 1991; Milanovic et al.; DeRemer; Smith). Egenskaperna är uppställda i en slumpmässig ordning oberoende av frekvensen i vilken de brukar framkomma i bedömarnas kommentarer.

- |  |                           |
|--|---------------------------|
| - grammatik                            | - rättskrivning           |
| - innehåll                             | - beaktande av            |
| - handstil                             | uppgiften/instruktionerna |
| - tempusformer                         | - disposition             |
| - skiljetecken och små/stora bokstäver | - koherens                |
| - problem med inledning                | - satsstrukturer          |
| - längd                                | - ordval                  |

I motsats till Tarnanens konstateranden om grammatikens dominans i bedömningsvärlden påpekar några forskare (t.ex. Cumming och Vaughan 1992 refererade i Weigle 2002) att det snarare är innehållet eller det första intrycket som bedömarna i högsta grad baserar sina val av nivå på i sina bedömningar. Å andra sidan rapporterar Tarnanen (2002) om studier (t.ex. Milanovic 1996) där påvisats att innehållet inte skulle spela någon roll vid bedömning. Med tanke på motstridiga resultat om ämnet anser jag att det är för komplicerat att fastställa vilka egenskaper i texter bedömarna betonar i allmänhet. I stället anser jag att man borde betrakta vilka egenskaper som betonas i vart och ett av de fall där bedömaren, testtagaren, kriterierna och uppgiften är givna och behandla bedömningar i förhållande till den kontext där bedömningen äger rum.

Eftersom denna studies forskningsfrågor (se kapitel 1) utgår från att undersöka hur bedömarna stödjer sig på nivåskalan är det särdeles intressant att få veta hur bedömare tolkar nivåbeskrivningar och om det finns brister i deras bruk av nivåskala. Tyvärr är det svårt att få tag på studier om ämnet om man betraktar just bedömning av skriftliga färdigheter och specifikt i svenska. Om man ser på forskningsresultat rent allmänt genom att överskåda studier som inriktats till bedömning i andra språk än svenska kan man dock få hänvisningar

till temat. Tarnanen (2002) exempelvis upptäckte att bedömarna ofta hade svårigheter med att använda nivåskalan. Enligt Tarnanen beklagade bedömarna sig över att den allmänna skalan var flertydig och obestämd, något som vållade problem i synnerhet i ändarna av skalan. Trots det rapporterar Tarnanen att bedömarna var relativt samstämmiga i sina bedömningar.

Utöver svårigheter med tillämpning av nivåbeskrivningar lyfter Tarnanen (2002:131) fram att bedömarna på basis av flera undersökningar tycks bedöma texter annorlunda än beskrivningarna i kriterierna anger. Detta anser forskaren vara en följd av t.ex. att nivåbeskrivningar är för abstrakta eller lämnar för mycket utrymme för bedömarens egna tolkningar, vilket förorsakar svårigheter med att tillämpa kriterierna i bedömningen (se avsnitt 2.2). Vidare påpekar Tarnanen att problem med att använda nivåskalan ändamålsenligt kan uppstå när bedömaren läser texten bara ytligt. Då riskerar bedömaren enligt forskaren att stödja sig på endast ett eller två kriterier i stället för hela nivåbeskrivningen.

Vaughan (1991) har upptäckt motsvarande företeelser som Tarnanen (2002). Även hon framhäver att bedömaren inte är någon maskin som kan tillämpa färdighetsskalan likformigt på varenda text som han/hon bedömer. Vaughan betonar att det är därför som flera bedömare ofta känner sig osäkra om kvaliteten av sina beslut och om deras bedömningar står i enighet med andra bedömares. Speciellt knepiga för bedömaren är enligt Vaughan gränsfallstexter som ofta innebär en stor obalans mellan olika språkliga egenskaper. Ett exempel på en sådan uppsats kunde vara en text som har ett livfullt och väl organiserat innehåll men bevisar ett bristfälligt ordförråd hos testtagaren. I dylika fall tenderar bedömaren ofta enligt Vaughan att strunta i kriterierna eller nivåbeskrivningarna och bedöma efter sin individuella bedömningsstil.

### **3 MATERIAL OCH METOD**

Syftet med detta kapitel är att redogöra för hurdan material och hurdan metod jag använde i min undersökning. I avsnitt 3.1 utreder jag hur jag kom till att använda mitt material. För detta syfte presenteras projektet inom vilket uppsatsen är insamlad samt redogörs kortfattat för en avhandling i vars material uppsatsen, bedömarna och nivåskalan var inkluderade. I avsnitt 3.2 redogör jag för nivåskalan och i avsnitt 3.3 för informanterna. Hur jag genomförde

det praktiska materialinsamlandet, förberedde mig för intervjuerna, intervjuade informanterna och transkriberade intervjuerna utreder jag i avsnitten 3.4 och 3.5. Analysmetoden behandlas i avsnitt 3.6.

### 3.1 Källan till materialet

Uppsatsen som bedömdes i denna undersökning är insamlad inom ett projekt kallat *På väg mot kommunikativ kompetens: tillägnandet av svenskans struktur hos finska inlärare*. Paavilainen et al. (2009) rapporterar att projektet är gemensamt för Åbo universitet och Jyväskylä universitet och har för avsikt att ”bidra till teoriutvecklingen inom språkinläringen och utvecklandet av svenskundervisningen för finska inlärare på högstadie- och gymnasienivå och vid universitet” (Paavilainen et al. 2009:144). Materialet vid projektet insamlades enligt författarna på högstadie-, gymnasie- och universitetsnivåer och används vid olika delprojekt av flera forskare. Eftersom det i denna studie används bara en uppsats skriven av en högstadieelev skall jag i det följande utesluta informationen om materialinsamlandet angående gymnasister eller universitetsstuderande.

Paavilainen et al. (2009) klargör att grundskoleprojektet var longitudinellt och materialet därför insamlades regelbundet i årskurserna 7-9. Materialinsamlingen skedde enligt författarna i två faser varav den första tidsmässigt hänförde sig till mellan våren 2006 när eleverna var sjundeklassare och våren 2008 när eleverna slutade i årskurs 9. Den andra fasen påbörjades våren 2007 och avslutades våren 2009<sup>1</sup>. Eleverna som deltog i projektet kom enligt Paavilainen et al. från fyra grundskolor i Södra och i Mellersta Finland. Var och en av eleverna hade finska som förstaspråk. Enligt författarna hade eleverna påbörjat sina studier i svenska i årskurs 7, vilket innebär att de läste svenska som B-språk. Vid grundskoleprojektet insamlades enligt Paavilainen et al. både skriftligt och muntligt material. Dessutom bestod det skriftliga materialet av uppsatser, test och kursprov.

Uppsatsen som jag använde i min undersökning hör till ett material som en studerande vid Jyväskylä universitet hade utvalt ur det material som insamlades vid projektet På väg mot kommunikativ kompetens. Studeranden använde materialet i sin pro gradu -avhandling, som vid tidpunkten då detta skrivs inte hade fullbordats. I studien ingick 220 uppsatser som

---

<sup>1</sup> en beräknad tidpunkt; då artikeln skrevs hade den andra fasen inte slutförts

bedömdes av fyra bedömare med hjälp av läroplanens nivåskala för språkkunskaper, och det är dessa bedömare och denna nivåskala som jag använde även i min undersökning.

Egentligen var det sex uppsatser som mina informanter bedömde i denna studie och som behandlades under intervjuerna men jag valde endast en uppsats vars bedömning jag i föreliggande undersökning analyserar. Bland de 220 uppsatserna som studeranden använde i sin pro gradu -avhandling valde jag ut delvis slumpmässigt och delvis med tanke på texternas kvalitet sex uppsatser som mina informanter bedömde ytterligare en gång i samband med denna undersökning. Meningen var att informanterna skulle bedöma de sex texterna på nytt utan att kontrollera nivåerna som de tidigare hade angivit för uppsatserna i samband med pro gradu -undersökningen. På detta sätt ville jag försäkra mig om att bedömningarna i denna undersökning skulle vara oberoende av de tidigare bedömningarna. I valet av de sex uppsatserna försökte jag ta hänsyn till att det skulle komma med flera olika texttyper för att få flera valmöjligheter för analysen. Dessutom ville jag att texterna inte skulle vara för korta för bedömningen, eftersom jag inte ville göra bedömningen slumpmässig eller mer komplicerad än den är i sig själv. Antalet texter begränsade jag på grund av det kvalitativa väsendet av denna studie samt på basis av det faktum att jag inte kunde förvänta mig att bedömarna skulle lägga för mycket tid på bedömningarna.

### **3.2 Nivåskala för språkkunskaper**

I denna undersökning använde jag således Utbildningsstyrelsens (25.8.2011) nivåskala för språkkunskaper som angår den grundläggande utbildningen i Finland. Enligt Utbildningsstyrelsen är skalan ”en i Finland utarbetad tillämpning av de skalor som ingår i den allmäneuropeiska referensramen för undervisning, inläring och bedömning av språk som utvecklats av Europarådet”. Nivåskalan för den grundläggande utbildningen innehåller nivåbeskrivningar för tio språkliga nivåer inom hörförståelse, tal, läsförståelse och skrivning. Bilaga 1 illustrerar hur en komprimerad version av nivåskalan för skriftliga färdigheter ser ut. Skalan i sin helhet kan ses på Utbildningsstyrelsens webbsida (Webb a).

Nivåskala för språkkunskaper är stigande på så sätt att nivå A1.1 beskriver språkliga färdigheter hos en nybörjare medan en elev på nivå C1.1 redan visar en relativt funktionell skriftlig färdighet. Själva bedömningen sker genom att jämföra nivåbeskrivningarna med

elevens prestation och på detta sätt placera eleven på en viss nivå (se avsnitt 2.2). Såsom framgår av skalan är nivåerna indelade i grupper. Till en grupp hör nivåerna vilkas symbol börjar på bokstav A, till en annan de med ett initialt B och till den tredje hör nivå C1.1. Dessutom är A-nivåerna indelade i A1- och A2-nivåer och på motsvarande sätt är B-nivåerna indelade i B1- och B2 -nivåer. Vidare har var och en av nivåerna försetts med ett namn (t.ex. ”De första grunderna i språket”) som kort beskriver de språkliga färdigheter som krävs på motsvarande nivå. Med formuleringen av skalan har man förmodligen syftat till att undvika ett lineärt intryck: avstånden mellan färdigheterna på olika nivåer är nödvändigtvis inte desamma (Tarnanen 2002).

### **3.3 Informanterna**

Som redan antytts var det fyra bedömare som utgjorde mina informanter i denna studie. Alla var kvinnliga och 38-57 år gamla. Tre av informanterna var finskspråkiga och en svenskspråkig. Alla de finskspråkiga informanterna hade avlagt de fördjupade studierna/laudatur-studierna i svenska vid ett universitet, och den svenskspråkiga informanten hade studerat lingvistik på svenska vid ett universitet i Sverige. Dessutom hade tre av informanterna, varav en var den svenskspråkiga, avlagt doktorsexamen och hade forskningserfarenhet. Alla fyra bodde i Finland vid tidpunkten då jag genomförde studien.

Var och en av de fyra informanterna var erfaren inom språkundervisning och språkbedömning. Alla rapporterade att de hade använt olika nivåskalor i bedömning av skriftliga färdigheter redan vid tidigare tillfällen, och även läroplanens nivåskala för språkkunskaper var bekant för dem från tidigare projekt. För att försäkra sig om att bedöma texter i linje med varandra hade informanterna dessutom vid tidigare projekt kommit överens om ett tiotal landmärken, dvs. uppsatser vars bedömning de tillsammans i förväg hade diskuterat.

För att skydda informanternas anonymitet skall jag tilldela bedömarna pseudonymer. Täcknamnen som jag skall använda är: Frida, Hanne, Kaisa och Laura. Med namnet Frida hänvisar jag till den svenskspråkiga informanten.

### 3.4 Det praktiska materialinsamlandet

Genomförandet av studien påbörjades med att kontakta informanterna. Jag skickade e-postmeddelanden till alla de fyra bedömarna i slutet av november och i början av december 2010 och frågade om de kunde tänka sig delta i min undersökning. I meddelandena presenterade jag kort att det var fråga om min kandidatavhandling samt syftet med studien. Alla svarade jakande. Efter att ha valt ut materialet som jag ville använda vid intervjuerna tog jag åter kontakt med informanterna per e-post för att komma överens om intervjutiderna och -platserna och för att skicka informanterna materialet. Detta bestod av ett följebrev med instruktionerna (bilaga 2), de sex uppsatserna som skulle bedömas, nivåskala för språkkunskaper (bilaga 1; se avsnitt 3.2), en bedömningsblankett (bilaga 3) samt en bakgrundsblankett (bilaga 4). Allt material var på finska eftersom även den svenskspråkiga informanten förklarade sig kunna förstå finska. I följebrevet bad jag informanterna att i förväg bekanta sig med uppsatserna och fundera över nivåer som de skulle ange för uppsatserna. På bedömningsblanketten kunde informanterna anteckna de tilldelade nivåer som de i denna studie kommit fram till samt vid behov de nivåer som de tyckte låg nära de tilldelade nivåerna. På blanketten fanns det också plats för bedömarens egna kommentarer. Bakgrundsblanketten följde löst den som Tarnanen (2002) använde i sin studie där hon studerade hur bedömare bedömer språkfärdigheter hos inlärare med finska som andraspråk. Blanketten använde jag för att få allmän kännedom om bedömarna och deras erfarenhet (se det föregående avsnittet).

Materialet skickade jag till tre av informanterna per e-post medan en av dem fick materialet i pappersform på sin arbetsplats. På grund av ett bristfälligt filnamn märkte två av dem som fick materialet per e-post inte att nivåskala för språkkunskaper fanns som bilaga, men de skaffade fram skalan på egen hand. Annars hade informanterna inga frågor eller uppgav inga oklarheter kring studien.

Tiden som bedömarna hade på sig att läsa och bedöma uppsatserna varierade från en vecka till flera veckor. Vid överenskomna tidpunkter intervjuade jag informanterna. De tre första intervjuerna ägde rum i mitten av december 2010 och den sista i början av januari 2011. Tre av bedömarna intervjuade jag på deras arbetsplatser och en på ett bibliotek. Alla fyra intervjuerna genomfördes emellertid på tu man hand i ett stängt rum. Intervjuerna varade i

cirka 45-55 minuter och spelades in på band med en mp3-spelare för senare analysering. De finskspråkiga informanterna intervjuade jag på finska och den svenskspråkiga på svenska.

Tre av informanterna, Hanne, Kaisa och Laura, hade fyllt i bakgrundsblanketten samt bedömningsblanketten på förhand. På bedömningsblanketten hade de tre antecknat även alternativa nivåer och kommentarer. Också på uppsatspappret hade Hanne, Kaisa och Laura antecknat de tilldelade nivåerna bredvid de motsvarande uppsatserna samt på ett par ställen gjort sina egna markeringar direkt i texter, såsom understrykningar. Dessutom gjorde Hanne och Laura några anteckningar under intervjun när de funderade över uppsatserna. Den informant som inte på förhand hade fyllt i blanketterna, Frida, besvarade frågorna på bakgrundsblanketten muntligt samt markerade de tilldelade nivåerna direkt på uppsatspappret bredvid de motsvarande uppsatserna under intervjun. Hon kommenterade och funderade också över alternativa nivåer muntligen samt gjorde flera markeringar i texter under intervjun.

### **3.5 Intervjuerna**

Den enklaste metoden för att få veta varför bedömaren bedömer uppsatser på ett visst sätt är förmodligen att fråga dem. Därför bestämde jag mig för att intervjua de fyra bedömarna om deras bedömningar. För intervjuerna förberedde jag mig genom att fundera på teman som jag ville ta upp med informanterna. Förebilden för intervjuerna var studien som Tarnanen (2002) genomförde i sin doktorsavhandling. Tarnanen använde en intressant intervjumetod som heter *halvstrukturerad temaintervju*. Metoden innebär att intervjuaren inte har några exakta intervjufrågor utan låter informanterna tala relativt fritt om det som de själva anser vara viktigt men som anknyter till de förutbestämda temana (se Hirsjärvi & Hurme 2000 och Kvale 1996). Det som intervjuaren kan göra i en halvstrukturerad intervju är att ställa specificerande frågor och försäkra sig om att diskussionen behandlar alla de teman som är nödvändiga. Jag anser att metoden lämpade sig bra för syftet med min studie: jag kunde inte styra informanterna för mycket åt ett visst håll och jag kunde inte med alltför strikta frågor förhindra dem att säga något som de ansåg vara viktigt, utan jag anser att informanterna själva lyfte fram det väsentligaste. Temana som jag emellertid ville ha behandlade följde delvis de teman som Tarnanen använde i sin avhandling. En av dem var helhetsintrycket, som i alla intervjuerna diskuterades först. Andra teman som i mindre eller större grad behandlades var:

- hur bedömaren började med bedömningen
- hur bedömaren tolkade kriterierna
- om det var lätt/svårt att placera uppsatserna på nivåerna
- hurdana problem det uppstod vid bedömningarna
- vilka kriterier som var avgörande för besluten
- om uppgiften var passande med beaktande av elevernas färdigheter

Till skillnad från Tarnanen i sin studie behandlade jag med mina informanter alla de sex uppsatserna så att bedömarna liksom tänkte högt när de läste texterna och motiverade varför de hade placerat texterna på vissa nivåer. Detta hjälpte mig med att få kännedom om hur bedömarna kommit fram till sina beslut och på vilka kriterier de mest grundade sina bedömningar.

Tack vare att alla fyra informanterna var pratsamma och öppna blev materialinsamlandet lätt och intervjuerna lyckade. Dessutom hade ingendera av informanterna kontrollerat bedömningsresultaten av de tidigare bedömningarna, även om några av dem rapporterade att de mindes texterna och att de skulle ha haft möjlighet till att kontrollera att de nu angav samma nivåer som de tidigare hade tilldelat samma uppsatser.

För att kunna analysera materialet transkriberade jag grovt intervjuerna, vilket innebär att jag inte markerade alla ljud eller upprepningar som hör till talspråket, men att jag försökte att inte förändra informanternas budskap. Pauser markerade jag oftast med en punkt, men det är inte alla pauser som jag prickade in. I några fall underströk jag ord eller stavelser för att markera betoningar. I dessa fall hade betoningar en funktion i att skapa betydelse.

### **3.6 Analysen**

För analysen valde jag ur de sex uppsatserna en uppsats som skulle bli målet för analyseringen. Mitt val grundade jag på att jag ville ha en uppsats som representerade ”en typisk uppsats för en högstadieelev” i anseende till dess längd, disposition, ordförråd, grammatisk felfrihet och de angivna nivåerna för att kunna betrakta hur informanterna bedömer en uppsats som i vissa avseenden förefaller vara normal i förhållande till elevgruppen. Jag ville nämligen inte välja en uppsats vars bedömning kunde ha varit

slumpartad på grund av att den var onödigt svår att bedöma, utan en uppsats som bättre skulle återspegla hur informanterna i allmänhet brukade bedöma. I valet av uppsatsen lade jag även märke till nivåerna som informanterna hade angivit. För att kunna studera skillnader i bedömningar ville jag välja en uppsats i vars bedömning informanterna inte hade varit helt enhetliga. På detta sätt kunde jag utesluta en uppsats. För att dessutom få bäst material för analysen beslöt jag mig för att välja en sådan uppsats som hade tilldelats flest nivåer. En uppsats som med hänsyn till mina kriterier visade sig vara lämpligast för syftet med denna studie var den första uppsatsen. Den hade tilldelats två identiska nivåbedömningar samt en nivå som var högre och en nivå som var lägre. Uppsatsen presenteras i avsnitt 4.1.

I det första steget av analysen försökte jag komprimera informanternas kommentarer om uppsatsen genom att stryka onödiga upprepningar och genom att skapa språkliga kategorier till vilka informanterna verkade hänvisa (se Kvale 1996; Tarnanen 2002). Teman som framkom under intervjuerna antecknade jag på fyra papper, ett för var och en av de informanterna. Med hjälp av anteckningarna försökte jag konstruera sammanfattande bedömarprofiler av informanterna genom att skriva upp det som jag ansåg vara det väsentligaste i att indikera hur informanterna hade kommit fram till vissa nivåer. I den andra fasen av analysen ville jag utreda hur bedömarnas beslut baserade sig på nivåskalan, och för detta syfte granskade jag noggrant nivåskalan och jämförde informanternas kommentarer med nivåbeskrivningarna med hjälp av de koncentrerade och kategoriserade bedömarprofilerna. Som sista steg återkom jag till transkriberingarna för att få en uppfattning om vilka kommentarer eller tankar som verkade vara avgörande för de slutgiltiga besluten som informanterna hade kommit fram till i sina bedömningar. I detta skede koncentrerade jag mig dessutom på likheter och skillnader mellan de fyra informanterna.

## **4 RESULTAT**

I detta kapitel presenteras resultaten av denna studie. Jag inleder med att i det följande avsnittet redogöra för den uppsats som bedömdes i denna undersökning, de angivna nivåer och de nivåbeskrivningar i nivåskalan för språkkunskaper som hänvisades till under intervjuerna. I avsnittet presenteras också de fyra informanternas bedömarprofiler som är ämnade att visa hur bedömarna motiverade sina val för de angivna nivåerna när de

intervjuades. I avsnitt 4.2 fortsätter jag med att redogöra för egenskaper i uppsatsen som informanterna tog upp under intervjuerna och i avsnitt 4.3 jämför jag dessa egenskaper med dem som står i nivåskalan. Likheter och skillnader mellan informanterna samt hur de eventuella skillnaderna påverkade bedömningsresultaten diskuteras i avsnitt 4.4.

#### 4.1 Hur uppsatsen bedömdes

Uppsatsen som bedömdes och vars bedömning jag analyserade presenteras i det följande:

Jag reser till Schweiz på sommaren med min familj. Vi tar båten från Helsingfors till Rostock, Tyskland. Därifrån åker vi med vår egen bil genom Tyskland till Schweiz. Där vi går till städer och museer. Det bästa var vy och Alper. Vi äter och dricker mycket. Sen åker vi tillbaka till Tyskland med bilen. Vi söker för ”Zimmer frei” men det var ingen. Vi stannade över natten i ett dyrt hotell. Sen vi tar båten från Rostock till Helsingfors och köper mycket Coca Cola och Fanta.

Rubriken på uppgiften var *Min resa* och det fanns inga inspirationsfrågor eller hjälpfrågor till eleverna i samband med uppgiften. I tabell 1 framkommer de angivna nivåerna.

Tabell 1 De angivna nivåerna

	<b>Frida</b>	<b>Hanne</b>	<b>Kaisa</b>	<b>Laura</b>
Bedömning	A1.3	A2.1	A2.1	A2.2

Såsom kan läsas i tabellen var bedömarna inte ense om nivån. Angivna blev tre nivåer som står närmast varandra i nivåskalan. Hanne och Kaisa tilldelade båda uppsatsen A2.1. Frida däremot angav nivån ett steg lägre, nivå A1.3, medan Laura tilldelade uppsatsen nivån ett steg högre, nivå A2.2.

I tabell 2 presenteras nivåbeskrivningarna för de nivåer som hänvisades till under intervjuerna (läroplansgrunderna 2004).

Tabell 2 Nivåbeskrivningarna som hänvisades till under intervjuerna

Nivå	Nivåbeskrivning
A1.2	<ul style="list-style-type: none"> <li>• Kan delge sina omedelbara behov med korta satser.</li> <li>• Kan skriva ett antal satser och fraser om sig själv och sin närmaste krets (t.ex. svar på frågor eller minneslappar).</li> <li>• Behärskar vissa grundläggande ord och uttryck och är förmögen att skriva mycket enkla huvudsatser.</li> <li>• Utantillärda fraser kan vara rätt skrivna men alla slags fel förekommer t.o.m. i synnerligen elementär fri skrivning.</li> </ul>
A1.3	<ul style="list-style-type: none"> <li>• Reder sig i enkla lätt förutsebara skrivuppgifter i de mest bekanta situationerna som anknyter till vardagliga behov och erfarenheter.</li> <li>• Kan skriva enkla texter (ett enkelt postkort, personuppgifter, enkel diktamen).</li> <li>• Behärskar de allra vanligaste orden och uttrycken som har att göra med detaljer eller konkreta behov i elevens eget liv. Kan skriva några meningar som består av enkla satser.</li> <li>• Många slags fel förekommer i elementär fri skrivning.</li> </ul>
A2.1	<ul style="list-style-type: none"> <li>• Reder sig i skrivuppgifter i de mest rutinmässiga vardagssituationerna.</li> <li>• Kan skriva korta, enkla meddelanden (personliga brev, lappar) som har att göra med vardagliga behov och enkla, katalogmässiga beskrivningar om mycket bekanta ämnen (verkliga eller fiktiva personer, händelser, egna och familjens planer).</li> <li>• Behärskar ett konkret ordförråd som anknyter till grundbehov samt grundläggande tempusformer och satser som är samordnade med enkla bindeord (och, men).</li> <li>• Kan skriva de allra enklaste orden och strukturerna ganska rätt men gör gång på gång fundamentala fel (tempusformer, böjning) och använder många klumpiga uttryck i fri skrivning.</li> </ul>
A2.2	<ul style="list-style-type: none"> <li>• Reder sig i skrivuppgifter i rutinmässiga vardagssituationer.</li> <li>• Kan skriva en mycket kort, enkel beskrivning av händelser, tidigare aktiviteter och personliga erfarenheter eller om alldagliga ting i sin livsmiljö (korta brev, minneslappar, ansökningar, telefonmeddelanden).</li> <li>• Behärskar det grundläggande alldagliga ordförrådet, strukturerna och de vanligaste sätten att binda ihop ord och satser.</li> <li>• Skriver enkla ord och strukturer rätt men gör fel i ovanligare strukturer och former och använder klumpiga uttryck.</li> </ul>

I det följande skall jag utreda hur informanterna motiverade sina val genom att presentera bedömarprofilerna av var och en av de fyra bedömarna. Avsikten med profilerna är att klargöra vilka inslag i uppsatsen informanterna funderade över under intervjuerna samt vilka egenskaper de enligt min tolkning tyckte var betydande för deras beslut. I redovisningen skall jag ta upp till diskussion enbart de kommentarer som direkt hänsyftar på den ifrågavarande uppsatsen och utesluta informanternas egna allmänna beskrivningar om deras sätt att bedöma eller hänvisningar till andra uppsatser.

Frida motiverade sitt val av nivå A1.3 genom att lista påfallande egenskaper i texten. Drag som hon lyfte fram var verbböjning och tempusformer, transfer, förståelighet, längd, sats- och meningsstrukturer och fel. Verbböjning och tempusformer verkade vara den avgörande faktorn i uppsatsen för att Frida inte ville tilldela texten en högre nivå, nivå A2.1. Hon bedömde texten att vara på nivå A1.3 just på grund av testtagarens bristfälliga användning av verbformer. I fråga om verbböjning och tempusformer blev Frida distraherad av att skribenten hade använt presens även om det i texten var fråga om en händelse i förfluten tid, och av att skribenten inte hade böjt verbet *ta*. Om sats- och meningsstrukturer kommenterade hon att texten innehöll endast huvudsatser, som dessutom var ganska enkla. Transfer syntes enligt informanten t.ex. i uttrycket *Zimmer frei* där informanten tyckte att skribenten hade påverkats av tyska. Frida refererade också direkt till nivåskalan. Inslag i uppsatsen som enligt henne stämde med skalan var att testtagaren visade sig behärska de allra vanligaste orden och uttrycken samt att testtagaren ”*reder sig i enkla lätt förutsebara skrivuppgifter i de mest bekanta situationerna som anknyter till vardagliga behov och erfarenheter*” (se tabell 2). Dessutom stödde Frida sig på skalan i det avseendet att texten var ganska enkel och att det förekom många slags fel i uppsatsen. Som textens starka sidor nämnde Frida förståelighet samt att texten var ganska lång.

Även Hanne tog upp testtagarens problem med tempusformer, men hon ansåg att textens förtjänster övervann bristerna. Det viktigaste för Hanne verkade vara att texten utformade en berättelse som var begriplig för läsaren. Andra goda kvaliteter i uppsatsen var enligt Hanne att texten innehöll en samordnad *men*-sats samt strävan efter att använda preteritum, även om den senare enligt informanten vållade påtagliga problem för testtagaren. Som exempel på inlärares försök till att använda preteritum nämnde Hanne ordet *köpade* som enligt informanten var ett bevis på att inlärares var medveten om hur man i allmänhet i svenska böjer ord i preteritum men just i detta fall inte visste den riktiga formen. Vidare påpekade Hanne att det var många verb som överhuvudtaget var böjda i texten, såsom *äter*, *dricker*, *åker* och *var*. En annan förtjänst i texten var enligt informanten att inlärares visade sig behärska det grundläggande ordförrådet i svenska. Dessutom fanns det några uttryck som informanten ansåg vittna om språkliga framsteg. Som exempel på sådana lyckade uttryck lyfte Hanne fram yttringarna *stannade över nattan* och *därifrån åker vi / sen åker vi*. Dessutom berömde informanten uppsatsen för satser med omvänd ordföljd. Liksom Frida tog

även Hanne upp att det fanns flera fel i uppsatsen men hon betonade att de inte förhindrade henne att förstå berättelsen.

Kaisa kommenterade först böjningsfel, som enligt henne uppträdde systematiskt i texten. Det som enligt informanten var bra var att testtagaren visade sig kunna yttra sig skriftligen på svenska. Vokabulären ansåg Kaisa vara tillräcklig för att texten skulle uppnå nivå A2.1. Informanten nämnde också några brister som stöd för sitt beslut, såsom bristfällig ordföljd och korta satser. Dessutom konstaterade informanten att testtagaren främst använde meningar med en enda sats med undantag av en samordnad *men*-sats. Till textens brister räknade Kaisa även påverkan av andra språk, vilket enligt henne blev synligt t.ex. i ordet *museums*. Nivåbeskrivningen på nivå A2.1 ”*kan skriva de allra enklaste orden och strukturerna ganska rätt men gör gång på gång fundamentala fel*” (se tabell 2) motsvarade enligt Kaisa uppsatsen träffande.

Laura berättade att hon baserade sitt beslut på textens omväxlande satsstrukturer. Detta motiverade hon med att skribenten inte inledde varenda mening med ordet *jag*, vilket enligt henne framgick bl.a. i uttrycket *därifrån åker vi med vor egen bil*. Dessutom tyckte Laura om de samordnade satserna med *men* och *och*. Laura ansåg att texten var förståelig och tillräckligt lång för att hon skulle kunna få en någorlunda uppfattning om skribentens ordförråd samt om hur skribenten bildade satser. De enstaka felen i orden *stadar* och *museums* samt influenser av tyskan tyckte informanten var så obetydliga att de inte förhindrade förståelsen. Det stora minuset i uppsatsen var enligt Laura bristerna i användningen av preteritumformer, vilket för henne var det mest distraherande draget. Problemen med tempusformerna framgick enligt Laura i att testtagaren hade använt presens även om det enligt informanten var uppenbart fråga om det förflutna. Det som informanten dock tyckte var ett bevis på att inläraren överhuvudtaget var medveten om olika tempusformer var att testtagaren hade använt uttrycket *vi stannade* i preteritum.

## 4.2 Vilka egenskaper som diskuterades

I tabell 3 försöker jag sammanfatta de egenskaper i uppsatsen som informanterna tog upp antingen som brister eller som styrkor. Jag har inte sorterat egenskaperna i positiva eller negativa kategorier eftersom informanternas bedömningar var enhetliga i fråga om olika

egenskapers kvalitet. Detta innebär att alla tyckte att t.ex. tempuset utgjorde ett minus och förståelighet ett plus för den angivna nivån. De egenskaper som informanten diskuterade under intervjun har jag markerat med en \* och de egenskaper som jag anser att informanten betonade mer än andra egenskaper eller som på något annat sätt verkade vara avgörande för resultatet har jag markerat med \*\*. Det bör märkas att kategorierna som jag har skapat inte är avsedda för att beskriva språket som sådant utan syftet är att utreda till vilka komponenter hos språkfärdighet mina informanter hänvisade under intervjun. Därmed hoppas jag att tabellen presenterar sådana faktorer i uppsatsen som informanterna ansåg att de angivna nivåerna bestod av. Kategorierna är emellertid inte absoluta i det avseendet att de skulle täcka allt som informanterna sade eller att de inte skulle överlappa varandra.

Tabell 3 Egenskaper i uppsatsen som diskuterades under intervjuerna

	<b>Frida</b>	<b>Hanne</b>	<b>Kaisa</b>	<b>Laura</b>
(Verb)böjning + tempus	**	*	*	**
Transfer	*		*	*
Förståelighet <sup>1</sup>	*	**	*	**
Uppsatsens längd	*			*
Fel	*	*	*	*
Ordförråd	*	*	*	*
Klarar sig i uppgiften/situationen	*			
Uppsatsens komplexitet/enkelhet	*			
Berättelse <sup>2</sup>		**		
Samordnade/underordnade satser	*	*	*	*
Ordföljd		*	*	
Satsernas längd/komplexitet	**		*	
Kan uttrycka sig skriftligen			*	
Omväxlande satsstrukturer/satsbildning		*		**

Tabellen visar att de egenskaper som alla fyra informanterna åtminstone i någon mån kommenterade var verbböjning och tempus, förståelighet, fel, vokabulär samt användning av

<sup>1</sup> Med ”förståelighet” menar jag i tabellen att ord, fraser, satser och meningar i sig själva är begripliga.

<sup>2</sup> I tabellen har jag kategoriserat Hannes intryck av att texten är en berättelse avskilt från kategorin ”förståelighet”. Man kan dock diskutera om begreppet berättelse mer skulle hänvisa till förståelighet.

samordnade eller underordnade sats. En annan frekvent egenskap i uppsatsen som diskuterades var transfer, som kommenterades under tre intervjuer. Ordföljd, längd på uppsatsen, satsernas längd och komplexitet samt omväxlande satsstrukturer/satsbildning diskuterades av två informanter. Påfallande är att skillnaderna mellan antalet egenskaper nämnda av var och en av de fyra bedömarna är obefintliga. Antalet faktorer som informanterna enligt tabellen baserade sina bedömningar på varierade mellan 8 och 10, då Frida nämnde 10 egenskaper, Kaisa 9 och både Hanne och Laura 8.

Det som däremot inte framgår av tabell 3 är i vilken grad bedömarna höll sig till nivåskala för språkkunskaper. Eftersom ett av syftena i denna studie är att utreda i vilken grad bedömarna använder nivåskalan och om detta påverkar bedömningsresultat skall jag i det följande avsnittet betrakta om de egenskaper som togs upp under intervjuerna egentligen var inkluderade i nivåskalan för språkkunskaper.

### 4.3 Hur bedömarna använde nivåskala för språkkunskaper

När man rannsakar nivåskala för språkkunskaper (se tabell 2) kan man märka att inte alla av de egenskaper som informanterna i denna studie ansåg vara viktiga för de angivna nivåerna egentligen uppträdde i skalan. I tabell 4 skall jag presentera dessa egenskaper samt vilka bedömare som hänvisade till dem<sup>1</sup>.

Tabell 4 Egenskaper som inte fanns med i nivåskalan men som togs upp under intervjuerna

	<b>Frida</b>	<b>Hanne</b>	<b>Kaisa</b>	<b>Laura</b>
Transfer	*		*	*
Förståelighet (Berättelse)	*	** (**)	*	**
Ordföljd		*	*	
Omväxlande satsstrukturer/satsbildning		*		**

<sup>1</sup> Det bör dock märkas att nivåskala för språkkunskaper är en holistisk skala som inte är avsedd att beskriva alla språkliga delområden uttömmande utan styra bedömaren att se det viktigaste i uppsatsen. Dessutom kan nivåbeskrivningarna i skalan sägas vara lite ungefärliga och allmänna, vilket betyder att de egenskaper som jag tycker inte är inkluderade i skalan kan någon annan tolka som att vara ett underbegrepp av ett sådant drag som står i skalan.

I tabellen står begreppet ”berättelse” inom parentes på grund av dess flertydighet. Intrycket av en berättelse är inte direkt inkluderat i nivåskala för språkkunskaper även om man skulle tolka begreppet ”berättelse” så att det hänvisar till textens förståelighet. I detta avseende verkar Hanne ha stött sig på sina egna kriterier i stället för de gemensamma. Å andra sidan kan uttrycket som Hanne använde under intervjun ”han/hon får berättelsen berättad” (se det följande avsnittet) anses hänsyfta på nivåskalan: nivåbeskrivningarna på nivå A2.1 ”*reder sig i skrivuppgifter i de mest rutinmässiga vardagssituationerna*” och ”*kan skriva – – enkla, katalogmässiga beskrivningar om mycket bekanta ämnen (verkliga eller fiktiva personer, händelser, egna och familjens planer)*” (se tabell 2) betyder verkligen i vissa avseenden att skribenten kan producera en berättelse. Det är alltså möjligt att Hannes kommentar om ”berättelse” vitt och brett hänvisade just till nivåbeskrivningarna på nivå A2.1 men att hon bara omtalade fenomenet i fråga på ett annat sätt.

Tabell 4 visar att alla av de fyra bedömarna stödde sig även på egenskaper i uppsatsen som inte fanns i nivåskalan. Hanne tycks ha tagit upp flest egenskaper i uppsatsen som inte tillhör nivåskalan med fyra sådana hänvisningar (om man inkluderar ”berättelse”), medan Frida verkar ha kommenterat minst sådana egenskaper med bara två referenser. Kaisa och Laura kommenterade om tre sådana inslag i uppsatsen som inte diskuteras i nivåskalan. När man beaktar även betoningarna, som jag har markerat med \*\* i tabell 4, verkar det vara Hanne och Laura som mest baserade sina beslut på sådana egenskaper i uppsatsen som inte tillhör nivåskalan.

Vidare visar tabell 4 att egenskaper som inte framträder i nivåskala för språkkunskaper men som trots det togs upp under intervjuerna var transfer, förståelighet, intrycket av en berättelse, ordföljd samt omväxlande satsstrukturer. Av dessa var förståeligheten den som mest påkallade kommentarer: var och en av informanterna stödde sig på förståeligheten. Även transfer, ordföljd och omväxlande satsstrukturer/satsbildning verkade vara populära val för motiveringar med tanke på de två till tre bedömarna som hänvisade till dessa kategorier under intervjuerna. Dessutom, såsom redan diskuterats, tog Hanne upp intrycket av en berättelse, vilket inte heller tydligt förekommer i nivåskalan.

#### 4.4 Vilka likheter och skillnader det uppstod

Frida var den som tilldelade uppsatsen den lägsta nivån, nivå A1.3. Hon funderade emellertid över att ange en högre nivå, nivå A2.1. Avgörande för att hon vid denna bedömningsgång inte angav nivå A2.1 var enligt hennes egna ord att skribenten inte kunde använda tempusformer eller böja verb. Följande diskussion visar hur Frida resonerade över om hon skulle tilldela nivå A2.1 i stället för A1.3. I replikerna står bokstav I för intervjuaren och F för Frida.

(1) I: Då borde man kunna använda tempusformer [för att kunna få nivå A2.1]?

F: Ja, grundläggande tempusformer – det funkar inte riktigt här.

Replikerna bevisar att tempusformerna var den begränsande faktorn i uppsatsen för Frida. Frida ansåg att uppsatsen var närmare nivå A2.1 än A1.2 men på grund av bristerna i tempusformer ville hon inte tilldela texten denna nivå utan kom fram till att ange en nivå mellan dessa två nivåer, nivå A1.3. Att Frida däremot inte tilldelade texten en lägre nivå, nivå A1.2 vållade inga problem. Där stödde Frida sig på nivåbeskrivningen ”*kan skriva några meningar som består av enkla satser*” på nivå A1.3 som enligt henne bättre motsvarade testtagarens färdigheter än nivåbeskrivningen ”*är förmögen att skriva mycket enkla huvudsatser*” på nivå A1.2. Replik 2 indikerar hur avgörande satsernas längd och komplexitet blev för Fridas beslut (uttrycken i kursiv stil är betonade):

(2) F: Jag tycker att det här med mycket enkla huvudsatser – ja på sätt och vis – men det är ändå *jag reser till Schweiz på sommaren med min familj*. Det tycker jag inte är *de allra mest enkla huvudsatser*. *Allra mest enkla huvudsatser* skulle ha varit jag reser till Schweiz.

Det väsentligaste i uppsatsen för Hanne i sin tur syntes vara att det blev en berättelse av texten. Tack vare uppsatsens begriplighet och flera förtjänster ville Hanne se genom fingrarna med tempusfelen och de andra bristerna och ange A2.1:

(3) H: Se mitä hän osaa ni hän saa sen tarinan kerrottua vaikka siellä on kaikenlaista virhettä. On kirjoitusvirhettä ja niin poispäin. Mutta kuitenkin.

*Det som han/hon kan är att han/hon får berättelsen berättad även om det finns alla slags fel där. Felskrivningar och så vidare. Men ändå.*

Hanne tycktes dela Fridas åsikt om tempusformerna i så måtto att även Hanne funderade över om texten med tanke på testtagarens problem med preteritum förtjänade nivå A2.1. Ett alternativ som enligt min åsikt kan ha förorsakat att bedömarna trots enigheten om tempusproblem bedömde uppsatsen olika är att bedömarna var oeniga om betydelsen av nivåbeskrivningarna. Kanske tolkade Frida nivåbeskrivningen *behärskar – – grundläggande tempusformer* så att inläraren som befinner sig på nivå A2.1 enligt nivåskalan måste behärska grundläggande tempusformer helt, och eftersom testtagaren visade sig ha betydande svårigheter med preteritum och verbböjning kom Frida fram till en lägre nivå. M.a.o. tyckte Frida kanske att nivåbeskrivningen på nivå A2.1 förutsätter att inläraren inte gör så många fel i tempusanvändningen som den ifrågavarande testtagaren hade gjort i uppsatsen, och kunde därför inte tilldela uppsatsen den högre nivån av de två möjliga. Hanne kan däremot ha tänkt att testtagaren inte nödvändigtvis behöver behärska alla de kunskaper som nämns i skalan på nivå A2.1 fullständigt utan om prestationen vittnar om stora styrkor kan bedömaren förbise bristerna i tempusformer och ändå ange nivå A2.1.

Det kan finnas även en annan förklaring till att Frida bedömde uppsatsen strängare än Hanne. Beaktansvärt är att Frida inte tålde skribentens brister på tempusformer även om det stod i nivåbeskrivningarna att testtagaren på nivå A2.1 kan göra ”*gång på gång fundamentala fel (tempusformer, böjning) och använd[a]er många klumpiga uttryck i fri skrivning*”. Kanske var det bara Hanne av dessa två bedömare som verkligen iakttog denna eftergift, och kunde därmed ange en högre nivå trots testtagarens problem med tempuset. I detta fall skulle det vara fråga om att Frida inte beaktade alla kriterier i nivåskalan, vilket förorsakade skillnaden mellan bedömningarna.

Att avgöra mellan nivåerna A1.3, A2.1 och A2.2 var lätt för Kaisa:

- (4) K: Mutta onha tässä nyt kuitenkin asiaa. Et osaa kirjottaa, osaa ilmaista itseään kirjallisesti. Ja tosiaan on sitä sanastookin. Ni mun mielestä se on ihan kakspisteykkösen verran.

*Men det finns ju många saker här. Liksom att kan skriva, kan uttrycka sig skriftligen. Och minsann innehåller det här även en rätt bra vokabulär. Så jag tycker att det räcker nog för tvåpunktett.*

Repliken visar att Kaisa, liksom Hanne, ville ta upp skribentens starka sidor i stället för att haka upp sig på fel eller brister. De mest avgörande styrkorna syntes vara att skribenten kunde skriva och uttrycka sig skriftligt, något som bara allmänt och mångtydigt hänvisades till i nivåskalan, samt ordförrådet. Även Kaisa övervägde rollen av tempusproblem i sin bedömning, men nivåbeskrivning för nivå A2.1 ”*kan skriva de allra enklaste orden och strukturerna ganska rätt*” motsvarade enligt henne uppsatsen så bra att hon trots fel i verbböjningen ville ange denna nivå. I likhet med Hanne men i motsats till Frida ansåg Kaisa kanske att uppsatsen innehöll så många element som tillhörde nivå A2.1 att hon kunde förbise tempusproblemen och tilldela texten denna nivå trots att den inte fullkomligt uppfyllde kraven på behärskande av grundläggande tempusformer. Möjligt är också att Kaisa, liksom Hanne möjligen, baserade sitt val på nivåbeskrivningen på nivå A2.1 ”*gör gång på gång fundamentala fel (tempusformer, böjning)*” även om hon, eller Hanne, aldrig nämnde denna nivåbeskrivning under intervjun. Kanske ansåg Kaisa och Hanne sålunda att testtagaren egentligen uppfyllde kraven på behärskning av tempusformer eftersom denna med beaktande av försöken till att använda preteritum uppenbarligen visste hur verb överhuvudtaget kan böjas, och att felen bara var sådana ”gång på gång fundamentala fel” i tempusformer och böjning som hänvisades till i nivåskalan. I alla fall verkar det vara antingen skillnaden i tolkningen av nivåbeskrivningarna eller Fridas förbiseende av en eftergift i nivåskalan som var den största orsaken till att Frida tilldelade uppsatsen en nivå som var lägre än den nivå som Hanne och Kaisa angav.

Laura tilldelade uppsatsen den högsta nivån. På samma sätt som Hanne och Kaisa men till skillnad från Frida betonade Laura textens förtjänster och uppfattade tempusfelen som mindre viktiga:

- (5) L: Jos tässä nyt kattoo sitä aikamuotoa niin sehän on tuolla aakaksykkösessä. Eli vois nyt kuvitella että onko imperfekti nyt semmonen sit joka pitäs jo osata tuolla. Mut mun mielestä tässä oli muuta hyvää joka tavallaan kompensoi sen imperfekti-osaamattomuuden. Eli siinä oli hyvää sanastoo ja ihan hyviä lauseita. Mutta verbit on vaikeita.

*Om man nu ser på tempuset så är den ju där på aatvåettan. Man kan alltså tänka sig om imperfektet nu är sådant som man redan borde kunna där. Men jag tycker att det fanns annat bra här som liksom kompenserar att imperfektet inte behärskas. Det fanns alltså bra vokabulär och rätt bra satser. Men verb är svåra.*

Påfallande i Lauras kommentar är att hon tyckte att uppsatsen på grund av dess brister i användning av preteritumformer utan vidare skulle ha kunnat befinna sig på nivå A2.1. De tre andra informanterna verkade ju vara mer misstänksamma om uppsatsen med beaktande av dess bristfälliga tempusformer kunde tilldelas nivå A2.1. Frida ansåg till och med att texten med tanke på dess tempusformer tillhörde en nivå lägre, nivå A1.3. Informanternas oenighet om tempuset bevisar att bedömarna inte bara betonade olika drag i uppsatsen i sina bedömningar utan också att de hade olika tolkningar av nivåbeskrivningar och deras betydelser.

I replik (5) tog Laura upp vokabulären och satser som motiveringen för sitt beslut, exempelvis att texten innehöll ”bra satser” som det inte direkt står någonting om i nivåskalan. Om man tolkar ”bra satser” på motsvarande sätt som Frida, som tyckte att uppsatsen inte innehöll bara ”de allra mest enkla huvudsatser”, kunde man anta att motiveringen grundade sig på nivåskalan. Senare under intervjun avslöjade Laura emellertid det som hon menade med ”bra satser” och som slutligen avgjorde den tilldelade nivån:

(6) L: Että ehkä lähinnä niiden vaihtelevien lauserakenteiden takia mä päätysin kyllä tohon aakakskakkosee.

*Kanske främst på grund av de omväxlande satsstrukturerna skulle jag nog komma fram till den där aatvåtvåan.*

Med omväxlande satsstrukturer avsåg Laura sådana som inte gärna för många gånger börjar med ordet *jag* utan som varierar på så sätt att de inleds med olika satsdelar. En sådan företeelse beskrivs emellertid inte i nivåskalan för språkkunskaper. I detta avseende verkar även Laura ha grundat sitt beslut på sina egna kriterier i stället för nivåbeskrivningarna. Dessutom, till skillnad från de tre andra bedömarna, ansåg Laura att testtagarens färdigheter uppnådde till och med nivå A2.2, något som de andra informanterna inte ens övervägde.

Därmed verkar Lauras mildhet i sin bedömning ha berott på att hon både ansåg att uppsatsens styrkor var fler och mer framträdande än de andra bedömarna tyckte och att felen i preteritumformer inte var så allvarliga. Därför ville hon förbise testtagarens brister och belöna uppsatsen genom att ange nivå A2.2 i stället för A2.1, som alltså med hänsyn till verbböjningen enligt henne skulle ha varit den mest passande nivån.

## 5 DISKUSSION

I de fyra första avsnitten av detta kapitel tar jag i tur och ordning upp mina fyra forskningsfrågor (se kapitel 1) och försöker med hjälp av resultaten i det föregående kapitlet besvara dem. I samband med svaren på forskningsfrågorna diskuterar jag även relationen i vilken mina resultat står till den forskningsbakgrund som diskuterades i avsnitt 2.4. I det sista avsnittet av detta kapitel, i avsnitt 5.5, reflekteras över tillförlitligheten och giltigheten av denna studie.

### 5.1 Forskningsfråga 1

Min första forskningsfråga löd: ”I vilken mån använder bedömarna nivåskala för språkkunskaper (Läroplansgrunderna 2004) när de bedömer uppsatsen?”. Denna studie visade att alla fyra bedömarna fäste avseende vid både egenskaper i uppsatsen som nämns i nivåskalan för språkkunskaper och egenskaper som inte nämns i nivåskalan. Det fanns emellertid skillnader rörande i vilken grad informanterna betonade olika egenskaper och om de egenskaper som saknades i nivåskalan var avgörande för informanternas bedömningar.

Att informanterna inte helt stödde sig på nivåskalan i sina bedömningar är inte överraskande i det avseendet att skalan som användes i undersökningen var en holistisk nivåskala som kan anses vara tämligen abstrakt och flertydig. Såsom diskuterades i kapitel 2 brukar holistiska nivåskalor kännetecknas av problemet att alla kriterier som borde vara med i bedömningen inte helt kan komprimeras i nivåbeskrivningarna (Tarnanen 2002). Därmed blir det ofta utrymme för bedömares egna tolkningar av betydelsen av nivåbeskrivningar, vilket kan leda till att bedömare bedömer texter på ett annat sätt än förväntats i nivåskalan. Att bedömarna

tolkar nivåskalan på olika sätt får däremot inte leda till subjektiva bedömningar utan de angivna nivåerna borde alltid grunda sig på den gemensamma skalan. Det är ju meningen att nivåskalan slår fast bedömningsgrunderna på samma bas oberoende av enskilda bedömares uppfattningar om vad som är en bra uppsats. I denna undersökning verkade var och en av de fyra informanterna i någon mån stödja sig på skalan, vilket antagligen garanterade en relativt hög grad av reliabilitet och validitet i bedömningsresultaten. Diskuteras kan dock om bedömningarna skulle ha varit mer överensstämmande om bedömarna i högre grad hade stött sig på nivåskalan.

Uppsatsen i denna undersökning var kanske svår att bedöma på grund av den obalans mellan olika språkliga färdigheter hos testtagaren som den omvittnade. Enligt informanterna kunde inläraren tydligen uttrycka sig skriftligt och hade ett relativt omfattande ordförråd. Problemet var att skribenten däremot inte verkade kunna böja verb ordentligt även om han/hon tydligen var medveten om olika tempusformer i svenska. Med anledning av obalansen i språkliga färdigheter ansåg några av bedömarna kanske att uppsatsen låg mellan två nivåer: mellan A1.3 och A2.1 enligt Frida och mellan A2.1 och A2.2 enligt Laura. Såsom diskuterades i kapitel 2 brukar det just vara vid sådana gränsfall som bedömarna mest tenderar att lita på sina egna tolkningar i stället för nivåskalan (Vaughan 1991), och det är kanske därför mina informanter i så hög grad baserade sina beslut på sina egna kriterier. Denna studie erbjuder emellertid ingen information om hur bedömarna skulle ha använt nivåskalan i bedömning av andra uppsatser eller om tendensen att använda egna kriterier bara var en tillfällighet.

## **5.2 Forskningsfråga 2**

Den andra forskningsfrågan var: ”Vilka kriterier använder bedömarna mest i sina bedömningar?”. I anseende till alla de möjliga faktorer som bedömarna skulle ha kunnat basera sina bedömningar på förefaller bedömarna i ljuset av denna undersökning ha varit mycket enhetliga i sina motiveringar. Informanterna syntes oftast kommentera egenskaperna i tempusformer och verbböjning, förståelighet, fel, vokabulär, samordnade/underordnade satser och transfer. Kategorierna sammanfaller med dem som diskuterades i avsnitt 2.4 med undantag av innehåll som i denna studie knappast fick någon kommentar. Resultatet ger stöd för teorierna enligt vilka uppsatsens innehåll inte skulle spela någon roll vid bedömning. Å

andra sidan kan frånvaron av hänvisningar till innehållet bero på att uppgiften i detta fall var mycket enkel och att den inte möjliggjorde närmare betraktande av innehållsmässiga drag.

Resultaten av denna studie motsvarar dem som presenteras i Tarnanen (2002) däri att grammatiska drag och vokabulär var de kategorier som mest kommenterades av alla fyra informanterna under intervjuerna. Liksom i Tarnanens undersökning hänvisades det även i denna studie bara allmänt till ordförrådet, medan grammatiken betraktades ur olika synpunkter. I motsats till Tarnanens studie avgjorde grammatiska drag emellertid enligt min tolkning den avgivna nivåbedömningen bara i Fridas fall, medan Hanne, Kaisa och Laura tycktes mer betona egenskaper som förståelighet och testtagarens sätt att bilda satser. Intressant nog var det således just den svenskspråkiga informanten som verkade vara mer grammatiskt orienterad. Diskuteras kan om informantens språkliga bakgrund påverkade hennes sätt att bedöma. I alla fall vågar jag med stöd av resultaten i denna studie påstå att åtminstone tre av mina fyra informanter inte var grammatiskt orienterade. Att bedömarna i stort sett inte baserade sina beslut bara på uppsatsens grammatiska fel eller grammatisk felfrihet var kanske en följd av att informanterna var väl utbildade och erfarna och därmed kunde beakta och evaluera även andra slags egenskaper i sina bedömningar.

Denna undersökning gav inte heller någon evidens för Charneys (refererad i Vaughan 1991) påstående om att bedömare skulle bruka iaktta bara ytliga egenskaper, såsom längd, felskrivningar, skiljetecken etc. Tvärtom omvittnar mina resultat att mina informanter inte baserade sina beslut på sådana ytliga drag i uppsatsen. Bedömarna i denna studie kommenterade några gånger textens längd eller stavfel men det förekom inte ett enda fall där informanten skulle ha lyft fram att hon baserade sitt nivåval på uppsatsens längd eller felskrivningar. Däremot konstaterade informanterna att dessa typer av kvaliteter och fel var av ringa vikt.

### **5.3 Forskningsfråga 3**

För det tredje ville jag utreda om det fanns kriterier som bedömarna använde i bedömningarna men som inte tillhörde nivåskalan. Kriterier som inte tillhörde nivåskala för språkkunskaper men som informanterna ofta stödde sig på i denna studie var förståelighet, transfer, ordföljd samt användning av omväxlande satsbildning och satsstrukturer, varav förståelighet

diskuterades under var och en av de fyra intervjuerna. De tre andra kriterierna fick kommentarer av tre eller två informanter. Det som är anmärkningsvärt i bedömarnas användning av nivåskalan är att det var så många av dem som motiverade sitt beslut med just förståelighet, transfer, ordföljd och omväxlande satsstrukturer/satsbildning även om dessa kriterier inte tydligt kom fram i nivåskalan för språkkunskaper. Kanske vittnar bedömarnas enhetlighet om en gemensam uppfattning om språket och om det som är viktigt i att kunna behärska i ett främmande språk, och kanske har denna gemensamma faktor inte kunnat sammanfattas till exakta ord i nivåskalan.

Att bedömare inte stödjer sig på nivåskalan utan använder sina egna kriterier – skapade antingen medvetet eller omedvetet - i stället för eller utöver de gemensamma kriterierna kan emellertid leda till att bedömningen blir osystematisk eller att några bedömare förefaller vara mildare och några strängare än de andra, vilket i värsta fall kan framkalla bedömarberoende mättningsfel i bedömningar (Tarnanen 2002). I denna undersökning kom det faktiskt fram att en bedömare bedömde uppsatsen strängare och en mildare än de två andra, men i beaktande av att informanterna bedömde bara en uppsats kan jag inte dra några slutsatser om informanternas stränghet eller mildhet i allmänhet. För detta borde man studera hur informanterna bedömer flera uppsatser.

#### **5.4 Forskningsfråga 4**

Sist undersökte jag om bedömarnas sätt att bedöma eller använda nivåskalan hade ett samband till de angivna nivåerna. Det visade sig i de föregående avsnitten att det fanns skiljaktigheter i hurdana egenskaper bedömarna uppmärksammade i uppsatsen, och några av dessa egenskaper framträdde inte ens i nivåskalan. Trots det fanns det ingen regelbundenhet i egenskaperna som bedömarna tog upp under intervjuerna eller i graden i vilken informanterna höll sig eller inte höll sig till nivåskalan som kunde förklara varför en av bedömarna bedömde uppsatsen mildare och en strängare än de två andra. Tvärtom framstod det att även den strängaste bedömaren kommenterade uppsatsens förtjänster och den mildaste textens brister, men ändå bedömde de uppsatsen olika. Avgörande för oenhetligheten gällande den mest passande nivån verkade däremot bli graden i vilken informanterna baserade sitt beslut på de faktorer som de kommenterade under intervjuerna och om de ansåg att dessa egenskaper var mer eller mindre tungt vägande för deras bedömning. Hanne och Laura kommenterade t.ex.

felen men tillade att de inte var så viktiga i förhållande till textens förtjänster. Laura som tilldelade uppsatsen den högsta nivån verkade kommentera samma brister som de tre andra informanterna, såsom tempusproblem, fel och transfer, men trots det kom hon fram till en nivå som var ett steg högre än den nivå som Hanne och Kaisa hade tilldelat uppsatsen och två steg högre än den som Frida hade angivit. På motsvarande sätt verkade Frida som tilldelade texten den lägsta nivån ta upp samma egenskaper i texten som de tre andra informanterna hade diskuterat, bland vilka även var positiva kommentarer om t.ex. textens förståelighet, längd, ordförråd samt användning av samordnade eller underordnade satser.

Att bedömarnas observationer om egenskaper i uppsatsen var likartade beror troligen på att informanterna var erfarna och välutbildade språkbedömare. Att de trots likadana observationer kom fram till att tilldela uppsatsen tre olika nivåer återspeglar däremot att de av någon anledning inte kunde tolka nivåbeskrivningarna och deras betydelse på samma sätt. Dessutom verkade tre av informanterna, de som tilldelade uppsatsen de högsta nivåerna, betona uppsatsens styrkor mer än den som angav den lägsta nivån. Vidare kunde oenigheten kring frågan om uppsatsen borde tilldelas nivå A1.3 eller A2.1 sammanhållas med att en av informanterna möjligen inte beaktade alla kriterier i nivåskalan.

Tarnanen (2002) talar om fenomenet haloeffekt när bedömaren inte uppmärksammar alla de faktorer som finns med i nivåskalan utan fattar beslut beroende av hans/hennes individuella uppfattningar och värdesättningar. Eftersom informanterna i denna studie ofta baserade sina val på egenskaper som inte tydligt hänvisades till i nivåskalan och inte kommenterade alla egenskaper som diskuterades i skalan kunde man tänka sig att informanterna misslyckades med att vara objektiva i sina bedömningar. Å andra sidan använde jag i denna studie en holistisk nivåskala vars funktion, såsom redan diskuterats, inte är att ange absoluta svar på vilka komponenter en bra uppsats består av utan att styra bedömaren åt att beakta vissa egenskaper i uppsatsen. Därmed kan det vara snarare nivåskalan än bedömarna som var själva problemet: alla de nödvändiga kriterier som man möjligen behöver vid bedömning av olika uppsatser har inte kunnat sammanfattas i nivåtabellen. Vidare kan det vara så att det faktum att några faktorer i skalan ofta togs upp medan andra aldrig hänvisades till har berott på uppgiften eller att uppsatsen helt enkelt inte erbjöd möjlighet att studera alla de egenskaper som beskrivs i nivåskalan. Nivåskala för språkkunskaper innehåller t.ex. beskrivningar som tar ställning till i vilka situationer testtagaren borde klara sig, men detta kan bedömaren inte avgöra på basis av en enda uppsats. Därmed anser jag att det med tanke på de begränsningar

och möjligheter som den holistiska nivåskalan medför skulle vara orimligt att förvänta sig att bedömarna hade kommenterat alla de egenskaper som finns i nivåskalan på ett likvärdigt sätt.

## 5.5 Om tillförlitligheten

Eftersom jag har studerat bedömning med kvalitativa metoder och baserat mina resultat på en analys av intervjuer, bör man ta i beaktande att jag inte har kunnat undvika att åtminstone i någon mån bli subjektiv. Även om jag har försökt studera ämnet objektivt är det alltid möjligt att jag har missuppfattat något som informanterna har sagt eller att mina resultat återspeglar mina egna uppfattningar eller ändamål. Ett exempel på en sådan företeelse som kunde vålla otillförlitlighet i studien är möjligheten att jag av misstag har styrt mina informanter åt ett visst håll och åt sådana kommentarer som inte annars hade dykt upp under intervjuerna, eller eventualiteten att jag har analyserat materialet så att resultaten stödjer min tankevärld i stället för att vara objektiva. Dessutom var informanterna inte likvärdiga i analysen på grund av att de hade förberedd sig för intervjuerna på olika sätt, vilket i detta fall innebar att en av informanterna fick kommentera uppsatserna extempore medan andra i förväg hade läst och funderat över uppsatsen och nivån de skulle ange.

Kritik kan också riktas till sambandet i vilket informanternas kommentarer stod till verkligheten. Det som informanterna sade under intervjuerna återspeglar deras egna uppfattningar om det som var viktigt i att ange nivåer men inte nödvändigtvis det som de tilldelade nivåerna i verkligheten var baserade på. Vidare kan man överväga möjligheten att informanterna omedvetet ville skapa ett positivt intryck på intervjuaren. Kanske var det så att bedömarna inte ville förefalla vara för grammatikorienterade och därför undvek kommentarer som hänvisade till grammatiska element i uppsatsen.

Reliabiliteten och validiteten lider också av det faktum att en av mina informanter även arbetade som min handledare för denna kandidatavhandling. För att skydda informantens anonymitet skall jag emellertid inte avslöja vem av de fyra informanterna den ifrågavarande bedömaren var. Det bör dock beaktas att hon kanske har besvarat mina frågor och annars agerat under intervjun på så sätt att hon skulle hjälpa mig med min avhandling i stället för att objektivt kommentera sina bedömningar, eller att jag, på grund av mina tidigare uppfattningar

om henne och med tanke på det framtida samarbetet med henne, har förvrängt hennes kommentarer eller resultaten i hennes fall.

Med hänsyn till de ovannämnda faktorerna kan reliabiliteten och validiteten i denna undersökning sägas vara tvivelaktiga, även om man skulle begränsa reliabiliteten och validiteten i samma grupp av informanter och i samma uppsats som i denna undersökning har använts. Uppenbart är att resultaten av denna studie inte kan generaliseras till fall där bedömarna och/eller uppsatsen är olika eller där det används flera uppsatser.

## 6 AVSLUTNING

I denna kandidatavhandling studerade jag hur fyra erfarna bedömare bedömde en svenskspråkig uppsats skriven av en finsk högstadieelev med hjälp av Utbildningsstyrelsens nivåskala för språkkunskaper. Uppsatsen valde jag så att den hade tilldelats tre skilda nivåer: A1.3 av en informant, A2.1 av två informanter samt A2.2 av en informant. Syftet var att studera i vilken mån bedömarna stödde sig på nivåskalan i sina bedömningar, vilka kriterier antingen i eller utanför nivåskalan bedömarna mest baserade sina bedömningar på samt varför de trots den gemensamma nivåskalan bedömde uppsatsen olika. För detta syfte intervjuade jag bedömarna och diskuterade med dem hur de bedömde uppsatsen och hur de motiverade nivåerna som de valde att tilldela uppsatsen. Metoden som jag använde vid intervjuerna var halvstrukturerad temaintervju. Resultaten fick jag genom att kvalitativt analysera intervjuerna.

Studien visade att alla fyra bedömarna i viss mån stödde sig på nivåskalan för språkkunskaper i sina bedömningar, men det var alla fyra som baserade sina beslut även på sådana kriterier som inte var med i den gemensamma skalan. Stora skiljaktigheter i graden i vilken bedömarna stödde sig eller inte stödde sig på nivåskalan kan inte urskiljas. Hur vägande för de angivna nivåerna användningen av bedömarnas egna kriterier i stället för eller utöver den gemensamma skalan var varierade emellertid i någon mån.

Trots oenigheten om nivån motiverade bedömare sina val med relativt likartade kriterier och kommentarer. Egenskaper i uppsatsen som mest diskuterades av bedömarna under intervjuerna var tempusformer och verbböjning, förståelighet, fel, vokabulär, användning av

samordnade och underordnade satser samt transfer. Av dessa var egenskaperna förståelighet, transfer och ordföljd sådana som inte tydligt tillhörde nivåbeskrivningarna i nivåskalan för språkkunskaper, varav förståelighet till och med diskuterades av var och en av de fyra bedömarna. Intressant i resultaten var att tre av de fyra bedömarna, i motsats till den allmänna uppfattningen om bedömarna och bedömning, inte visade sig grunda sina val på graden av grammatisk felfrihet utan mer betonade egenskaper som förståelighet och omväxlande satsstrukturer i uppsatsen. Uppsatsens innehåll verkade inte heller spela någon avgörande roll för bedömningarna i denna undersökning.

Oenigheten om den angivna nivån sammanhänge med skillnader mellan bedömarnas sätt att tolka nivåbeskrivningarna. Bedömarna verkade också i varierande grad betona antingen uppsatsens styrkor eller brister på så sätt att de som lade vikt vid uppsatsens starka sidor ville ange en högre nivå än den som framhävde bristfälligheter i texten. Dessutom berodde det faktum att en informant ville tilldela uppsatsen en lägre nivå eventuellt på att hon möjligen inte tog i beaktande alla kriterier i nivåskalan utan förbisåg en eftergift i nivåbeskrivningarna.

Resultaten av denna studie kan inte generaliseras till fall där dessa fyra bedömare skulle bedöma en annan uppsats eller flera uppsatser, för att inte tala om till fall där andra bedömare skulle bedöma denna uppsats, en annan uppsats eller flera uppsatser. Därför erbjuder studien kunskap bara om hur de fyra informanterna bedömde den uppsats som användes i denna studie. Det som föreliggande studie däremot kan erbjuda är en inblick i den stora variationen i motiveringar och val som bedömare möjligen i allmänhet baserar sina bedömningar på, samt i hur dessa motiveringar och val kan påverka bedömningsresultaten.

Kunskapen om bedömningsprocess och faktorer som samverkar i bedömning har en stor betydelse i utbildningen av bedömare och i skapandet av nivåskalorna, och det är inte minst bedömarna själva som har nytta av kännedomen om bedömningsteori. Med hjälp av teoretiska kunskaper kan bedömarna reflektera över sina val och styra sitt agerande och på så sätt sträva efter mer objektiva bedömningar. Därför behövs det flera undersökningar om hur bedömare i allmänhet bedömer olika slags uppsatser samt om frågan huruvida deras sätt att bedöma texter påverkar deras benägenhet att bedöma texter mildare eller strängare än andra bedömare. Med olika undersökningsmetoder kunde man studera skilda aspekter kring språkbedömning och få svar på olika slags forskningsfrågor. Intressant vore t.ex. att jämföra kvalitativa resultat med kvantitativa genom att exempelvis räkna medeltal av angivna nivåer för en stor mängd

uppsatser där varje uppsats skulle ha bedömts av flera bedömare. Med kvantitativa metoder kunde man också studera enighet mellan bedömare eller konsekvens hos enskilda bedömare och genom att sammankoppla resultaten av dessa beräkningar med analysering av intervjuer utreda om bedömningsmotiveringar uppvisar ett samband till enigheten eller konsekvensen. Värt att undersöka vore också bl.a. det som bedömare själva anser om nivåskala för språkkunskaper, eller hur bedömare själva kommenterar sitt sätt att bedöma.

## LITTERATUR

- Council of Europe. 2009. *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning* [online]. Strasbourg: Skolverket. Enheten för moderna språk. [Tillgänglig 25.8.2011]  
[http://www.skolverket.se/2.3894/publicerat/2.5006?\\_xurl\\_=http%3A%2F%2Fwww4.skolverket.se%3A8080%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2Fblob%2Fpdf2144.pdf%3Fk%3D2144](http://www.skolverket.se/2.3894/publicerat/2.5006?_xurl_=http%3A%2F%2Fwww4.skolverket.se%3A8080%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2Fblob%2Fpdf2144.pdf%3Fk%3D2144).
- Hirsjärvi, S. & Hurme, H. 2000. *Tutkimushaastattelu: teemahaastattelun teoria ja käytäntö*. Helsingfors: Helsinki University Press.
- Huhta, A. & Takala, S. 1999. Kielitaidon arviointi. I: Sajavaara, K. & Piirainen-Marsh, A. (red.), *Kielenoppimisen kysymyksiä*, Jyväskylä: Centret för tillämpad lingvistik. Jyväskylä universitet, 179-228.
- Kvale, S. 1996. *Interviews: an introduction to qualitative research interviewing*. Thousand oaks: Sage Publications.
- Lindroos, S. 2010. *Bedömningen av muntliga språkkunskaper i svenska : en analys av bedömarnas reflektioner kring elevbedömningens reliabilitet*. Helsingfors: Helsingfors universitet, humanistiska fakulteten, institutionen för nordiska språk och nordisk litteratur.
- Paavilainen, M., Järvinen E-L, Åberg A-M, Lahtinen, S., Rahkonen, M. 2009. Vägar att undersöka svenskan hos finska inlärare – datainsamling i ett forskningsprojekt. I: Berg, K. M. et al. (red.), *En färd i språket. Festskrift till Marketta Sundman på 60-års dagen den 12 mars 2009* [online]. Åbo: Åbo universitet, Nordisk filologi, 144-158. [Tillgänglig 25.8.2011] <http://www.doria.fi/bitstream/handle/10024/43995/festskrift-sundman.pdf?sequence=1>.
- Tarnanen, M., Huhta, A. & Pohjala, K. 2007. Mitä on osaaminen? Kielitaidon arviointi vastaajana. I: Luukka, M-R & Pöyhönen, S. (red.), *Kohti tulevaisuuden kielikoulutusta. Kielikoulutuspoliittisen projektin loppuraportti*, Jyväskylä: Centret för tillämpad lingvistik. Jyväskylä universitet, 381-412.
- Tarnanen, M. 2003. Kielitaidon arviointi: valintoja ja neuvottelua. I: Nissilä, L., Vaarala, H. & Martin, M. (red.), *Suolla suomea. Perustietoa maahanmuuttajien suomen kielen opettajille*. Äidinkielen opettajain liiton vuosikirja XLVII, Helsingfors: Äidinkielen opettajain liitto, 143-160.
- Tarnanen, M. 2002. *Arvioija valokeilassa: suomi toisena kielenä –kirjoittamisen arviointia*. Jyväskylä: Centret för tillämpad lingvistik. Jyväskylä universitet.

Utbildningsstyrelsen. Grunderna för läroplanen för den grundläggande utbildningen 2004

[online]. [Tillgänglig 25.8.2011]

<http://www02.oph.fi/svenska/ops/grundskola/LPgrundl.pdf>.

Vaughan, C. 1991. Holistic assessment: What goes on in the rater's mind? I: Hamp-Lyons, L.

(red.), *Assessing second language writing in academic contexts*, New Jersey: Ablex Publishing Corporation, 111-126.

Webb a. Nivåskala för språkkunskaper [online]. [Tillgänglig 25.8.2011]

<http://www.oph.fi/svenska/ops/grundskola/BILAGA2.doc>.

Weigle, S.C. 2002. *Assessing writing*. Cambridge: Cambridge University Press.

# BILAGOR

Bilaga 1 (fortsätter på nästa sida): Nivåskala för språkkunskaper för skriftliga färdigheter

A1.1	A1.2	A1.3	A2.1	A2.2	B1.1	B1.2	B2.1	B2.2
Kielländans alketiden hallinta	Kehittyvä alketiskiel- taito	Tommyva alketiskiel- taito	Pertuskiel- taidon alkuvaihe	Kehittyvä pertuskiel- taito	Tommyva pertuskiel- taito	Sujuva pertuskiel- taito	Isentäisen kielländän pertuseno	Tommyva isentäinen kiellianto
* Osa joutkon erillisiä sanoja ja sanoiltoja.	* Osa joutkon perussanoja ja sanoiltoja ja pystyy kirjoittamaan hyvin yksinkertaisia päälauseria.	* Osa kaikkein tavallisimpia sanoja ja ilmauksia, jotka liittyvät omaan elämään tai konkreettisiin tapisiin. Osa kirjoittaa muutamia yksiläuseria virheitä.	* Osa käyttää peruskäsitteisiin liittyviä konkreettisia sanoja ja perussanoiltoja sekä yksinkertaisia sidosseinoja (ja, mutta) liitettyjä rinnastaisia lausetta.	* Osa arktisen perussanoiltoja, rakenteita ja tavallisimmat sidosseinoja.	* Osa useimpien tutuissa ilmauksissa tarvittavien tekstien laadintaan riittävän sanoiltoja ja rakenteita, vaikka tekstissä esiintyy interferenssiä ja ilmeisiä kiertämissanoja.	* Hallitsee melko monimutkaisen kirjoittamiseen tarvittavaa sanoiltoja ja lauserkäsitteitä. Osa ilmauksia rinnastaisuuksia ja alisteruuksia.	* Osa laajan sanoiltoja ja vaativia lauserkäsitteitä sekä kielläiset keinoiltoja. sidosseinojen tekstien laadinnaksi. Sävyn ja tyylin joutavuus on riittävä ja pitkässä esityksessä voi ilmeisiä hyppäyksiä asasta toiseen.	* Kiehellinen ilmausvarasto ei riitä havaittavasti kirjoittamista.
* Ei kykene vaparaasen, tuotokseen, mutta kirjoittaa oikein muutamia sanoja ja ilmauksia.	* Ulkoa opetellut fraasit voivat olla oikein kirjoitettuja, mutta alkeellimmassakin vaparaassa tuotokseen esiintyy hyvin paljon kaikenlaisia virheitä.	* Alkeellimmassakin vaparaassa tuotokseen esiintyy momenttisia virheitä.	* Kirjoittaa kaikkein yksinkertaisimmat sanat ja rakenteita melko oikein, mutta tekee toistuvasti virheitä perussanoiltoja (sidosseinoja, tärvtus) ja tuottaa paljon kiertämissanoja vaparaassa tuotokseen.	* Kirjoittaa yksinkertaiset sanat ja rakenteita oikein, mutta tekee virheitä harvinaisimmassa rakenteissa ja muodissa ja tuottaa kiertämissanoja.	* Rutiinomainen kielländän ja perusrakenteita ovat jo suhteellisen virheettömiä, mutta joutuu vaativammia rakenteita sanoiltoja tuottavai ongelma.	* Pystyy kirjoittamaan yhtämittaisia ja kohtuullisen virheettömiä kiellä, vaikka virheitä esiintyy vaativissa rakenteissa, tekstin jäsentäessä ja tyylissä ja vaikkakin äidinkielen tai jonkin muun kielen vaikutus on ilmeinen.	* Hallitsee melko hyvin oikeinkirjoituksen kielländän ja välimerkkien käytön eräitä virheitä joutuu väärinkäytöksiin, tyylissä ja ilmausten ja tyylin joutavuus tuottavai ongelma.	* Hallitsee hyvin kielländän, sanoiltoja ja tekstin jäsentäksen. Virheitä voi esiintyä harvinaisissa rakenteissa ja ilmauksissa sekä tyylisäntöissä.



## Bilaga 2: Följebrev

### **Hyvä arvioija!**

Kiitos, että osallistut kandidaatintutkielmaani liittyvään tutkimukseen. Ohessa toimitan arvioitavat tekstit (6 kpl), arviointilomakkeen, taitotasokuvaukset sekä taustatietolomakkeen.

Pyydän, että luet arvioitavat kirjoitelmat haluamassasi järjestyksessä haluamallasi tavalla ja mietit, mille taitotasolle asettaisit kunkin kirjoitelman käyttäen taitotasokuvauksia. Merkitse taitotaso (esim. A1.1) arviointitaulukkoon oikean tekstin kohdalle. Voit merkitä kohtaan ”vaihtoehtoinen taso” tason, joka mielestäsi seuraavaksi parhaiten sopisi kyseiselle tekstille, jos vaihtoehtoinen taso on kovin lähellä valitsemaasi tasoa. ”Mahdolliset kommentit” – kohtaan voit merkitä huomioitasi esim. tekstistä, kriteereistä, käyttämästäsi arviointistrategiasta tai tekstin kirjoittajasta. Olisin iloinen, jos myös yleisellä tasolla kirjoittaisit muutamalla sanalla ylös jotain arvioinnistasi, esimerkiksi kuinka arviointisi eteni, missä järjestyksessä ja kuinka olet päätenyt arvioihisi. Näin haastattelussa asiat muistuvat ehkä paremmin mieleen. Voit kirjata havaintojasi esim. arviointilomakkeen kääntöpuolelle.

Pyydän myös, että lyhyesti vastaat taustatietolomakkeen kysymyksiin. Lomakkeen tarkoituksena on suurpiirteisesti kartoittaa arvioijan taustaa, eikä tietoja analysoida sen tarkemmin.

Kaikki tiedot niin haastattelussa kuin lomakkeissakin käsitellään luottamuksellisesti. Jos sinulla on kysyttävää, vastaan mielelläni kysymyksiin. Kiitoksia yhteistyöstäsi!

Ystävällisin terveisin,  
Miia Juutilainen

## Bilaga 3: Bedömningsblankett

<i>tekstinro</i>	<i>taso</i>	<i>vaihtoehtoi nen taso</i>	<i>mahdolliset kommentit</i>
1			
2			
3			
4			
5			
6			

Bilaga 4: Bakgrundsblankett

## Arvioijien taustatietolomake

Nimi

---

Ikä \_\_\_\_\_

Äidinkieli \_\_\_\_\_

Koulutus

---

Ruotsin opinnot

---

Millaista kokemusta sinulla on ruotsin kielestä (esim. tutkimuskokemusta, opetuskokemusta jne.)? Kuinka pitkältä ajalta ja miltä tasolta?

---

---

Kuinka paljon olet arvioinut kirjoittamista?

---

---

Oletko käyttänyt aiemmin taitotasokuvauksia arvioinnissasi?

---

**Kiitos ajastasi!**