

Inburgeringexamens in Nederland: op zoek naar de wetenschappelijke achtergrond. Een literatuurstudie.

David Koenen

Studentnummer 3853802

MA Taal, mens en maatschappij

Begeleider: dr. Jacomine Nortier

Tweede lezer: dr. Carolien van den Hazelkamp

28 augustus 2018

Samenvatting

In deze scriptie staan de inburgeringsexamens lees-, schrijf-, spreek-, en luistervaardigheid centraal. Er wordt getracht aan de hand van een literatuurstudie een antwoord te vinden op de mate waarin de inburgeringsexamens wetenschappelijk gefundeerd zijn.

In het eerste gedeelte van deze scriptie wordt een overzicht gegeven van de vier verschillende taaldeelvaardigheden: spreken en schrijven (productieve taalvaardigheid) en luisteren en lezen (receptieve taalvaardigheid). Verreweg de meeste aandacht zal uitgaan naar de deelvaardigheden spreken en luisteren, omdat die voor inburgeraars het meest relevant zijn.

In het tweede gedeelte wordt een overzicht gegeven van de inburgeringsprocedure in Nederland, en de rol die taal daarin speelt. Ook wordt kort stilgestaan bij de inburgeringsprocedure in een select aantal andere landen, met speciale aandacht voor de verschillen tussen Nederland en België.

Hierna wordt aandacht besteed aan de toetsen die gebruikt worden in de Nederlandse inburgering. Omdat de wetenschappelijk verantwoording en de inhoud van de inburgeringsexamens van het verantwoordelijk ministerie niet openbaar gemaakt mogen worden, kunnen er hierover echter geen sluitende uitspraken gedaan worden. In plaats daarvan zullen de oefentoetsen, die op internet vrij beschikbaar zijn, besproken worden.

Ten slotte worden de oefentoetsen spreek- en luistervaardigheid afgezet tegen toetsingscriteria uit de literatuur, zodat bekeken kan worden in hoeverre die toetsen overeenstemmen met wetenschappelijke bevindingen. Concluderend kan gesteld worden dat de inburgeringstoetsen gedeeltelijk overeenstemmen met bevindingen uit de literatuur, maar zouden kunnen profiteren van een meer communicatieve insteek, die nu grotendeels ontbreekt.

Table of Contents

1	Introductie	4
2	Methode	5
3	Theoretisch kader	6
3.1	Geschiedenis en achtergrond	6
3.2	Leesvaardigheid	8
3.3	Schrijfvaardigheid	10
3.4	Spreekvaardigheid	13
3.5	Luistervaardigheid	18
4	Inburgering in de praktijk	23
4.1	De inburgeringsprocedure in Nederland	23
4.1.1	De geschiedenis van inburgering in Nederland	25
4.2	De rol van taal in de inburgeringsprocedure: taaltoetsing en taalanalyse	26
4.2.1	De rol van taal in toetsing en de getoetste taalvaardigheidsniveaus	26
4.2.2	LADO: taalanalyse in de asielprocedure	27
4.3	De internationale context: de inburgeringsprocedure in andere landen	28
4.3.1	Hoe gaan andere landen om met inburgering?	28
4.3.2	Verschillen tussen de inburgeringsprocedures van Nederland en Vlaanderen	29
4.4	Toetsing in de Nederlandse inburgeringsprocedure	30
4.4.1	De in de Nederlandse inburgeringsprocedure gebruikte toetsen	30
4.4.2	Examenonderdeel leesvaardigheid	31
4.4.3	Examenonderdeel schrijfvaardigheid	32
4.4.4	Examenonderdeel spreekvaardigheid	34
4.4.5	Examenonderdeel luistervaardigheid	35
5	Discussie	37
6	Conclusie	42
7	Bibliografie	44

1 Introductie

Het afgelopen decennium werd gekenmerkt door een grote golf van massaprotesten, opstanden, revoluties, en burgeroorlogen in de Arabische wereld. Deze zogenoemde Arabische Lente begon bijna acht jaar geleden, op 18 december 2010, met opstanden in Tunesië als reactie op de publieke zelfverbranding van een straathandelaar, wiens handelswaar door de politie in beslag was genomen.

Hierop volgden protesten in onder meer Algerije, Jordanië, Oman en Irak. In Tunesië, Egypte en Jemen werden regeringen omvergeworpen. In de meeste van deze landen werd het na verloop van tijd weer relatief rustig. Soedan gaat nog altijd gebukt onder gewelddadige aanslagen, en in Jemen, Libië, Syrië, Afghanistan, en Irak woeden nog altijd hevige burgeroorlogen.

De effecten van het voorduren van de oorlogen in Afrika en het Midden-Oosten zijn ook in Nederland nog altijd te merken. De vluchtelingenstroom werd in 2015 zo groot dat de Immigratie- en Naturalisatiedienst (IND) in maart van dat jaar bekendmaakte 100 extra medewerkers in dienst te nemen om het sterk toegenomen aantal asiel- en nareisverzoeken tijdig te kunnen behandelen.¹ Hiermee werd het medewerkersbestand verdubbeld.

Aan het verkrijgen van een verblijfsvergunning in Nederland zijn eisen verbonden. Vluchtelingen die al in Nederland zijn wanneer zij een asielaanvraag doen, moeten een inburgeringsproces doorlopen voordat zij een definitieve verblijfsvergunning toegewezen kunnen krijgen. Gezinsleden voor wie een nareisverzoek wordt gedaan, moeten ditzelfde proces al in het land van herkomst doorlopen en aantonen dat zij beschikken over basiskennis van de Nederlandse taal en maatschappij.

Een van de belangrijkste onderdelen van het inburgeringsexamen is het afleggen van vier taalvaardigheidsexamens (leesvaardigheid, schrijfvaardigheid, spreekvaardigheid, en luistervaardigheid) voor het Nederlands. Deze examens staan in deze scriptie centraal. De centrale onderzoeksvraag hierbij is in hoeverre deze examens in de praktijk gefundeerd zijn op theoretische en/of empirische bevindingen uit de wetenschap. Een tweede onderzoeksvraag is of er in de literatuur aanwijzingen te vinden zijn die zonodig zouden kunnen zorgen voor verbetering van deze examens.

¹ *IND zet extra mensen in voor groeiend aantal nareisaanvragen*, nieuwsbericht in NRC Handelsblad op 30 maart 2015. Dit artikel is te vinden via deze link: <https://www.nrc.nl/nieuws/2015/03/30/ind-zet-extra-mensen-in-voor-groeiend-aantal-nareisaanvragen-a1417481>

2 Methode

Om deze onderzoeksvragen te beantwoorden zullen zowel de wetenschappelijke als de praktische component afzonderlijk besproken worden in de vorm van een literatuurstudie. Hoofdstuk 3 van deze scriptie omvat een overzicht van recente wetenschappelijk literatuur op het gebied van de methodes van taaltoetsing (al dan niet in een immigratie- en/of tweedetaalonderwijscontext), en het bepalen van de validiteit van taaltoetsen.

Vervolgens wordt in hoofdstuk 4 van deze scriptie de praktische kant van dit onderwerp behandelen. Allereerst wordt kort het verloop van het inburgeringsprocedure in Nederland beschreven. Hierna komt de rol van taal in de asiel- en inburgeringsprocedure aan bod, waarbij ook kort aandacht zal zijn voor de rol van taalanalyse in asielzaken (daarbij gaat het om analyse van specifieke taalkenmerken waarmee getracht wordt na te gaan of iemand echt uit een geclaimde plaats, land, of streek afkomstig is). Vervolgens wordt globaal weergegeven hoe enkele andere landen met het inburgeringsproces omgaan, en hoe dit verschilt met de gang van zaken in Nederland. Hierbij wordt ook aandacht besteed aan internationale afspraken op dit gebied.

Hierna wordt gedetailleerder ingegaan op de taaltoetsen die in de Nederlandse inburgeringsprocedure gebruikt worden. Het gaat er hierbij om een beeld te krijgen van de soorten taalvaardigheid (lezen, schrijven, spreken, en luisteren) die getoetst worden, en op welke manier die precies getoetst worden. Ook wordt aandacht besteed aan de wetenschappelijke die daaraan ten grondslag liggen.

In hoofdstuk 5 worden de bevindingen uit de eerste twee delen tegen elkaar afgezet om te weten te komen of de taaltoetsing in de inburgeringspraktijk voortvloeit uit wetenschappelijke bevindingen, en zo ja, in hoeverre dat het geval is. Ook wordt er gekeken of er in de besproken literatuur informatie te vinden is die – indien nodig – mogelijkheden biedt tot verbetering van deze toetsen.

3 Theoretisch kader

In dit hoofdstuk wordt allereerst een kort overzicht gegeven van de geschiedenis van het onderzoeksgebied van taaltoetsing. Hierna wordt recente literatuur op het gebied van taaltoetsing behandeld. Omdat het inburgeringsexamen bestaat uit aparte onderdelen die er elk op gericht zijn een discreet deel van de taalvaardigheid te toetsen, zal ook in dit hoofdstuk de informatie over de toetsing van leesvaardigheid, schrijfvaardigheid, spreekvaardigheid, en luistervaardigheid in aparte secties behandeld worden. Op deze manier is het gemakkelijker om later per onderdeel te evalueren in hoeverre de op dit moment gebruikte toetsing overeenkomt met recente wetenschappelijk bevindingen.

3.1 Geschiedenis en achtergrond

Taaltoetsing is al lange tijd een populair gebied van onderzoek binnen de taalkunde. In Bachman (2000) wordt de geschiedenis van het onderzoeksgebied in de laatste twee decennia van de vorige eeuw beschreven. Voor een nog verder teruggaand achtergrondoverzicht wordt verwezen naar Spolsky (1995) en Barnwell (1996).

In de jaren '60 en '70 van de vorige eeuw werd taaltoetsing hoofdzakelijk op twee manieren benaderd. In het taaltoetsingsonderzoek was de *unitary trait hypothesis* van Oller (1979) leidend. Deze theorie ging ervan uit dat het menselijk taalvermogen één algemene vaardigheid was.

Aan de andere kant had in de taaltoetsingspraktijk de zogenoemde *skills and components view* (Lado, 1961) de overhand. Die benadering keek naar het menselijk taalvermogen als een verzameling vaardigheden (lezen, schrijven, spreken, luisteren) en componenten (zoals grammatica en vocabulaire). Bij het toetsen van taalvaardigheid was men hier steeds gericht op één discreet element van taal (Bachman 2000, p. 2-3). Volgens Bachman (1990, p. 82) was het *skills and components*-model beperkt in die zin dat het geen rekening hield met de bredere context waarin taal wordt gebruikt (zoals de specifieke situatie waarin een taalgebruiker zich bevindt, maar ook interactie tussen de vaardigheden en componenten).

Vanaf de jaren '80 werd taalvaardigheid steeds meer gedefinieerd als een interactief, dynamisch, communicatief proces. Binnen deze communicatieve benadering kreeg gespreksvaardigheid (een combinatie van de deelvaardigheden spreken en verstaan) een centrale plaats als graadmeter voor taalvaardigheid (Kwakernaak, 2011). Ook werd er steeds meer onderzoek gedaan op het gebied van tweedetaalverwerving. Dit had zijn weerslag op het onderzoek naar taaltoetsing, omdat er steeds meer factoren ontdekt werden die invloed zouden kunnen hebben op

taaltoetsen (Bachman 1990, p. 4). In Pienemann et al. (1988) werd een beroep gedaan op onderzoekers om voortaan ook het proces van taalontwikkeling mee te nemen in toetsontwerp en –beoordeling.

In de jaren '90 werd eveneens veel vooruitgang geboekt, op meerdere gebieden. Zo deed de domeingerichte scoreinterpretatie zijn intrede. Dit is een manier om een toetsscore te interpreteren op basis van inhoudelijke criteria, bijvoorbeeld de eis dat iemand een formulier moet kunnen invullen of om informatie moet kunnen vragen. Deze manier van interpreteren staat tegenover normgerichte scoreinterpretatie, waarbij individuele scores worden afgezet tegen de scores van een normgroep (Bossers et al. 2010, pp. 368-370). Hiernaast werd *item response theory* (IRT) een van de meest gebruikte modellen om toetsscores te analyseren (Bachman 2000, p. 5; Bachman & Eignor 1997). Naast de kwantitatieve benadering van taaltoetsing werd nu ook een kwalitatieve benadering populairder, waarbij onder meer de effecten van de kenmerken van toetsdeelnemers en de strategieën die zij gebruiken werden onderzocht (Banerjee & Luoma, 1997).

Op praktisch gebied kende dit decennium misschien wel de grootste verandering. Met behulp van de computer werd grootschalige toetsafname mogelijk. Ook maakte de computer het mogelijk toetsen steeds beter af te stemmen op individuen. Wel was nog onbekend in hoeverre deze ontwikkeling de validiteit van toetsen zou beïnvloeden (Bachman 2000, p. 9).

Naast deze ontwikkelingen werd er meer onderzoek gedaan naar factoren die toetsperformance beïnvloeden, zoals de strategieën die deelnemers gebruiken wanneer ze een toets maken, toets- en beoordelaarseigenschappen, en de eigenschappen van deelnemers zelf (p. 11).

Een opvallende ontwikkeling in de jaren '80 en '90 was de belangstelling voor taaltoetsing vanuit ethisch perspectief. Het gaat hierbij om een benadering van toetsing die zich vooral richt op de effecten van toetsen op sociaal, politiek, en institutioneel gebied. Het begrip *washback* (de mate van invloed die toetsing heeft op instructie, docent, en leerling (Bossers et al. 2010, p. 371)) vindt hier bijvoorbeeld zijn oorsprong. Maar vanuit deze hoek werd ook kritiek geuit op de mogelijke *gatekeeping*-functie van taaltoetsen (Spolsky, 1997). Proponenten van *critical social theory* zoals Shohamy (1997) gaan zelfs zo ver te zeggen dat het werkelijke doel van taaltoetsing vaak meer politiek is dan wordt voorgesteld. (Dit aspect van taaltoetsing zal in het hier volgende niet meer aan bod komen.)

Een kernbegrip in de literatuur over taaltoetsing is *validiteit*. De validiteit van een toets wordt bepaald aan de hand van de afweging of daadwerkelijk die vaardigheden en kennis gemeten worden als waar de toets voor ontworpen is. Het is van belang dat een toets zoveel mogelijk meet van het toetsonderwerp, maar niet meer dan dat. Een leesvaardigheidstoets moet bijvoorbeeld zoveel

mogelijk de leesvaardigheid toetsen, en niet ook de luistervaardigheid. Dit aspect van validiteit heet *inhoudsvaliditeit*. Ook wanneer dit gewaarborgd is, kan het voorkomen dat toetsresultaten geen accurate weerspiegeling geven van de vaardigheden van een cursist. Dit kan het gevolg zijn van een lage *constructvaliditeit*, die optreedt wanneer bijvoorbeeld de vorm van de vraagstelling niet aansluit bij de te meten vaardigheid. Overigens kan een niet passende toetsvorm ook bij de deelnemende cursisten tot verzet leiden. Wanneer zij een toets niet passend vinden en ertegen in verzet komen, is er sprake van *indruksvaliditeit* (Bossers et al. 2010, pp. 361-362).

Naast valide moet een toets ook *betrouwbaar* zijn. De betrouwbaarheid van een toets wordt bepaald door de mate waarin de resultaten een goed beeld geven van de vaardigheden van de deelnemers, zodat die duidelijk onderscheiden kunnen worden. Vragen die door iedere cursist goed (of fout) worden beantwoord, bieden geen mogelijkheid tot onderscheiding op vaardigheids- en kennisniveau, en dragen dus ook niet bij aan het opbouwen van een betrouwbaar beeld van vaardigheden en kennis van de verschillende deelnemers (Bossers et al. 2010, p. 363). Validiteit en betrouwbaarheid zijn complementair: een hoge mate van betrouwbaarheid is een voorwaarde voor een hoge mate van validiteit (Bachman 1990, p. 160).

In het hier volgende zal per deelgebied van taalvaardigheid – dus lees-, schrijf-, spreek-, en luistervaardigheid, zoals die in het inburgeringsexamen getoetst worden – een kort overzicht gegeven worden van een selectie van recente literatuur. Er zal uitgebreider worden ingegaan op de onderdelen spreek- en luistervaardigheid, aangezien die het meest relevant zijn voor inburgeraars.

De literatuur over taaltoetsing overlapt vaak met andere onderzoeksgebieden, zoals die van de psychologie. Dit zorgt ervoor dat de theoretische inhoud van de literatuur vaak erg complex en zeer uitgebreid is. Om deze reden zal de focus in dit hoofdstuk liggen op empirisch onderzochte toetsmethoden en hun relatie met validiteit. Dit zal ook de latere analyse van de inburgeringsexamens vergemakkelijken.

3.2 Leesvaardigheid

Sinds de communicatieve benadering binnen het taaltoetsingsonderzoek sinds midden jaren '70 van de vorige eeuw aan populariteit begon te winnen, is er binnen het tweedetaalonderwijs steeds meer aandacht ontstaan voor het gebruik van de doeltaal als voertaal. Kwakernaak (2011, p. 22) noemt het gebruik van de doeltaal als voertaal een ideaal binnen het tweedetaalonderwijs.

Hoewel dit een ideaal mag zijn, blijkt uit de literatuur dat het gebruik van de moedertaal van leerlingen wel degelijk kan helpen bij het leren van een tweede taal. Yamashita (2002) stelt dat, in

navolging van de *linguistic interdependence hypothesis*, aangenomen kan worden dat leesvaardigheid in de L1 de leesvaardigheid in de T2 positief beïnvloedt. De auteur stelt tevens dat volgens de *linguistic threshold hypothesis* aangenomen mag worden dat een zekere hoeveelheid van T2-vaardigheid nodig is voordat de T1-vaardigheden hun effect op het T2-leerproces kunnen hebben (deze hypothese is dus niet alleen van toepassing op de leesvaardigheid).

Vergelijkend literatuuronderzoek van Alderson (1984) op dit gebied liet zien dat T1-leesvaardigheid en T2-vaardigheid beide een effect hebben op T2-leesvaardigheid. Hierbij moet worden aangetekend dat de mate van T2-taalvaardigheid in deze een sterker effect had bij leerlingen met een lager T2-vaardigheidsniveau.

Shohamy (1984) vond dat deelnemers aan T2-leesvaardigheidstoetsen significant betere resultaten behaalden wanneer zij toetsitems gepresenteerd kregen in hun T1, ten opzichte van toetsen die helemaal in de doeltaal afgenomen werden. Ook hier bleken deelnemers met een algemeen lager T2-niveau een groter effect te ondervinden van de toetsing in hun T1 dan deelnemers met een hoger vaardigheidsniveau.

Volgens Alderson (2000) voegt het toetsen in de T2-doeltaal een extra graad van moeilijkheid aan een leesvaardigheidstoets toe. Hierdoor wordt het moeilijker vast te stellen of de mate van begrip die een T2-leerder heeft afhankelijk is van de opdracht, of dat die beïnvloed wordt door de daadwerkelijke taalvaardigheid in de T2.

Waar onderzoek voorheen vooral werd uitgevoerd op basis van open vragen, voerde Rahimi (2007) meer recent een studie uit waarin ditzelfde gegeven werd getoetst met behulp van meerkeuzevragen. Hieruit kwamen vergelijkbare resultaten naar voren: over het algemeen leek het niet uit te maken of de toetsitems in de T1 of T2 werden gepresenteerd, maar deelnemers met een laag T2-vaardigheidsniveau presteerden significant beter wanneer de toets in de T1 werd afgenomen. Toetsitems in de T1 verminderde daarnaast het spanningsniveau van deze groep deelnemers, omdat zij de toets als makkelijker ervoeren. Al met al vergroot dit de validiteit van leesvaardigheidstoetsing voor in elk geval de groep met een lager taalvaardigheidsniveau, omdat de invloed van het toetsontwerp op de prestaties wordt verminderd, zodat docenten ervan uit kunnen gaan dat hun beoordeling accurater is. Dit maakt eventuele beslissingen die op basis van de resultaten van zo'n toets genomen worden ook rechtvaardiger (p. 163).

Recent onderzoek in Nederland op dit gebied werd gedaan door Thomassen & Hacquebord (2017). Zij legden Nederlandse T2-Engels-leerlingen op niveau A2 en B1, opgedeeld in sterke, gemiddelde, en zwakke leerlingen, twee toetsen voor met dezelfde vragen, waarbij in één versie de vragen in het Engels (de doeltaal) waren geformuleerd (p. 14).

Ook hier werd een soortgelijk resultaat gevonden als in de hierboven besproken studies: de voertaal maakte gemiddeld gezien geen verschil, zowel op niveau A1 als B1. Alleen bij de groep zwakke leerlingen bleek dat zij de toets relatief beter maakten wanneer die in de T1 werd afgenomen. Alleen op het B1-niveau was dit verschil significant. De auteurs hadden dit effect eerder bij de A2-groep verwacht. Mogelijk is dit een resultaat van het kleine aantal deelnemers (175) aan het onderzoek. Anderzijds presteren de sterkere leerlingen beter op de Engelse toetsvariant, al is dit niet significant. Concluderend stellen de auteurs dat toetsing in de doeltaal mogelijk moet zijn op zowel A2- als B1-niveau, al merken zij hierbij dat de vraagstelling in de toetsen bewust op of beneden het taalniveau van de toets was geformuleerd (p. 19).

Het gebruik van doeltaal of voertaal in T2-leesvaardigheidstoetsen heeft dus mogelijk een groot effect. Vraagstelling in de T1 van zwakke T2-leerlingen kan ervoor zorgen dat hun T2-begrip minder negatief beïnvloed wordt, wat een positief effect heeft op de validiteit van de toetsen. Voor een praktische toepassing van dit principe op grotere schaal, is het natuurlijk wel noodzakelijk dat de groep deelnemers dezelfde T1 hebben. Bij kleinere groepen met verschillende moedertalen kan dit lastiger zijn, of in ieder geval een grote hoeveelheid werk met zich meebrengen.

3.3 Schrijfvaardigheid

Net als binnen het onderzoeksgebied van de leesvaardigheid, richtte schrijfvaardigheidsonderzoek zich gedurende de afgelopen decennia op de manier waarop schrijfvaardigheid zich ontwikkelt. Ook dit onderzoeksgebied, en de ontwikkelingstheorieën die hieruit zijn voortgekomen, is in hoge mate verweven met de psychologie.

Camp (2012) geeft een overzicht van de theorieën van schrijfvaardigheidsontwikkeling die in de 50 jaar voor publicatie zijn ontwikkeld. Een voorbeeld van zo'n theorie is de *syntactic maturity*-hypothese (Hunt, 1965), die stelt dat schrijfvaardigheidsontwikkeling is af te meten met behulp van het concept *T-unit*. Een T-unit wordt gedefinieerd als de kleinste mogelijke grammaticaal correcte eenheid in een regel tekst (p. 300). Hoe groter deze eenheid, des te groter is de syntactic maturity van de producent. Deze ontwikkeling is afhankelijk van de algemene cognitieve ontwikkeling (Hunt, 1970). Toch is syntactic maturity ook gesignaleerd zonder dat het gepaard ging met een vooruitgang in schrijfvaardigheid, wat twijfels deed ontstaan over de waarde van de theorie voor het beoordelen van schrijfvaardigheid (Camp 2012, p. 95).

Later ontstond meer twijfel over het koppelen van cognitieve ontwikkeling aan de ontwikkeling van schrijfvaardigheid. Er ontstonden twijfels bij de vraag of een analyse van

schrijfvaardigheid een accuraat beeld kon geven van de cognitieve ontwikkeling van een persoon. De omstandigheden waarin een schrijfofdracht gemaakt wordt, kunnen deze bijvoorbeeld mogelijk negatief beïnvloeden. Daarnaast is het mogelijk voor een schrijfofdracht te oefenen, wat eveneens een vertekend beeld zou kunnen schetsen van iemands cognitieve ontwikkeling (p. 96).

Uit haar analyse van psychologische theorieën van schrijfvaardigheidsontwikkeling concludeert Camp dat er vooral meer duidelijkheid moet komen over de manieren waarop ontwikkelingstheorieën de beoordeling van schrijfvaardigheid (mogelijk) beïnvloeden: "Different theories of development take assessments in different directions, influencing what materials are solicited, what criteria are applied, in what context the assessment will occur, what parties will play a role in assessing, what claims will be possible from the data, and what purpose(s) the assessment will serve." (p. 101).

Toch lijken, volgens de auteur, de meeste ontwikkelingstheorieën een aantal aannames te delen die kunnen dienen tot het ontwikkelen van een meer uniforme manier om schrijfvaardigheid te beoordelen. Een van deze aannames is de notie dat ontwikkeling contextueel is en waardesystemen en sociale conventies reflecteert van de maatschappij waarin een taalgebruiker leeft. Ook de aanname dat ontwikkeling multi-dimensioneel is komt naar voren. Dit betekent dat toetsing van discrete facetten van taalontwikkeling onvoldoende is om een goed beeld van de ontwikkeling te krijgen, net als het generaliseren over taalontwikkeling op basis van momentopnamen (een zeer beperkt aantal toetsen). Hiernaast zijn onderzoekers het erover eens dat de motivatie van een taalleerder een zeer sterke voorspeller is voor taalontwikkeling, en dat toetsprestaties niet altijd gelijk te stellen zijn met de mate van ontwikkeling (p. 101).

Op praktisch vlak is de afgelopen jaren onder Nederlandse onderzoekers een interessante discussie ontstaan over het gebruik van meerkeuzevragen als toetsingsmiddel voor schrijfvaardigheid. Bossers et al. (2010, p. 361) zijn stellig: schrijfvaardigheid kun je niet meten met behulp van meerkeuzevragen, omdat de vorm van de vraagstelling niet aansluit bij de te meten vaardigheid, wat leidt tot een lage constructvaliditeit.

Pullens et al. (2013) delen deze gedachte. Meerkeuzetoetsing zou leiden tot problematiek bij de beoordeling (p. 32) omdat ontwikkelingsvoortgang hiermee nauwelijks zichtbaar gemaakt wordt. Ook kunnen meerkeuzetoetsen voor schrijfvaardigheid geen rekening houden met schrijversvariabiliteit. Hierbij moet gedacht worden aan voorkeuren voor bepaalde schrijfofdrachten, of verschillen in achtergrondkennis (Stokking, 2000). Deze variabiliteit komt bij toetsing met open vragen vaker, maar ook niet altijd, naar voren.

Meerkeuzetoetsing zou wellicht wel een oplossing kunnen zijn voor problemen bij de beoordeling van toetsen, bijvoorbeeld omdat het corrigeren van meerkeuzeopdrachten een stuk simpeler gaat dan het beoordelen van opstellen, waarbij de beoordelingscriteria onduidelijk geformuleerd kunnen zijn. Ook beoordelaarseffecten zoals sequentie-effecten zouden hiermee tenietgedaan kunnen worden (Pullens et al. 2013, p. 32).

Volgens de auteurs zijn problemen met schrijversvariabiliteit en beoordelaarsvariabiliteit het kleinst bij meerkeuzetoetsing. Deze vorm van toetsing is relatief objectief, maar brengt wel het nadeel met zich mee dat het hiermee alleen mogelijk is om schrijfvaardigheid indirect te toetsen (p. 33). Omdat productieve schrijfoopdrachten ook andere facetten van taalvaardigheid aanspreken, menen de auteurs dat productieve schrijfoopdrachten en receptieve meerkeuzetoetsing verschillende kanten van de schrijfvaardigheid toetsen (p. 35). Maar volgens de auteurs is het onduidelijk welke soort schrijfvaardigheid meerkeuzetoetsen precies meten, wat wijst op een lage constructvaliditeit. Om deze reden raden zij het gebruik van meerkeuzetoetsen af (p. 40).

Dit onderzoek werd beantwoord door De Glopper & Willemsen (2014), die de conclusie van Pullens et al. (2013) overhaast en onvoldoende gefundeerd noemen. De Glopper & Willemsen (2014) stellen de volgende definitie van schrijfvaardigheid voor: "schrijfvaardigheid is een eigenschap die wij aan personen toeschrijven op grond van verschillen in de kwaliteit waarmee zij problemen in de schriftelijke communicatie oplossen" (p. 32).

Productie van geschreven tekst lijkt dus inderdaad een voorwaarde voor het kunnen beoordelen van schrijfvaardigheid. De auteurs brengen hier enkele argumenten tegenin. Allereerst menen zij dat de kennis die iemand moet aanspreken om een tekst te produceren, ook aangesproken worden bij toetsing met gesloten vragen. Ten tweede zijn niet alle verschillende vaardigheden die bij het schrijven van een tekst komen kijken, bij iedereen even goed ontwikkeld. Deze sterke en zwakke punten zouden bij gesloten toetsing beter afzonderlijk zichtbaar worden dan bij toetsing met open schrijfoopdrachten (p. 33). Om die reden zou toetsing van schrijfvaardigheid met open vragen voor diagnostische doeleinden ontoereikend zijn (de auteurs noemen beoordelaarseffecten bij open toetsing als een andere reden), en zou gesloten vraagstelling wel degelijk ingezet kunnen worden om deelaspecten van schrijfvaardigheid te toetsen (p. 37).

Enkele bezwaren tegen recente ontwikkelingen op het gebied van schrijfvaardigheidstoetsing worden aangedragen in Hamp-Lyons (2014). Zo zou geautomatiseerde beoordeling van geschreven teksten lijden onder een enorme toename van toetssystemen, onder meer onder invloed van de snelle ontwikkeling van corpusanalysetechnologieën (p. A1). De literatuur rondom deze technologieën zou weinig inzicht bieden in de manier waarop de geautomatiseerde beoordeling

werkt; het wordt alleen duidelijk dat menselijke niveau van beoordeling benaderd wordt (p. A2). Tevens zou automatisering de weg openen naar steeds grootschaligere toetsing, wat de afstand tussen leerling en docent zou vergroten (p. A3)

3.4 Spreekvaardigheid

Om te bepalen hoe spreekvaardigheid het beste getoetst kan worden, is het van belang te weten wat spreekvaardigheid precies is. In de literatuur wordt deze vraag op verschillende manieren beantwoord, al zijn er ook veel overeenkomsten aan te wijzen.

Fulcher (2014) maakt, alvorens te proberen spreekvaardigheid te definiëren, een voorbehoud: “No operational construct definition can ever capture the richness of what happens in a process as complex as human communication, even if the speaking test is mediated by tape or computer. Nor is it being suggested that a test of speaking should test everything that we will attempt to put into our list. Choices will depend upon test purpose and the extent to which scores need to be generalisable to other contexts.” (p. 19).

Anders dan Lado (1961), die spreekvaardigheid binnen zijn *trait theory approach* benaderde als een puur talige kwestie, stelt Fulcher dat er ook rekening gehouden moet worden met contextuele factoren die het spreekproces en de toetsing daarvan beïnvloeden. Hoewel Fulcher een verschil tussen competence en performance erkent, kan er volgens hem geen duidelijk onderscheid tussen de twee gemaakt worden omdat er altijd sprake is van interactie. In het geval van tweedetaalleerders zal de spreekperformance volgens hem afhangen van het geheel aan begrip, (cognitieve) vaardigheden, taalvaardigheid, de context waarin gesproken wordt, en de gesprekspartner(s) (p. 25).

Aan de hand van deze begrippen is het echter niet mogelijk te definiëren wat nu precies kenmerken zijn van spreekvaardigheid. Fulcher bekijkt spreekvaardigheid zowel als productie van spraakgeluid als het gebruiken van gespreksstrategieën. Op het niveau van spraakproductie is uitspraak, zeker voor tweedetaalleerders, vaak lastig. Een nieuwe taal kan bijvoorbeeld klanken bevatten die in de moedertaal van de leerder niet voorkomen. Dit kan ervoor zorgen dat de leerders moeite hebben om deze klanken te onderscheiden. Een voorbeeld uit het Engels is het verschil tussen de klinkers in de woorden ‘rule’ en ‘put’ (p. 25). Dit verschil is subtiel, maar wel degelijk aanwezig. Een onzorgvuldige uitspraak zorgt in dit geval echter niet voor betekenisverschil.

Tweedetaalleerders van het Nederlands kunnen hier echter wel mee te maken krijgen. De Nederlandse ‘ui’ is bijvoorbeeld vaak lastig voor NT2-leerlingen. Wanneer iemand die klank in het

woord 'huis' niet goed kan uitspreken en in plaats daarvan 'hoes' produceert, treedt er ook betekenisverschil op omdat 'huis' en 'hoes' in het Nederlands een *minimal pair* zijn: twee woorden die slechts in één spraakklank verschillen, waarbij dat verschil ook een betekenisverschil met zich meedraagt. Hierbij moet aangetekend worden dat een luisteraar meestal wel uit de context kan afleiden wat er bedoeld wordt, zodat eventuele verwarring minimaal zal blijven. De vraag is dus hoeveel gewicht dergelijke fouten in spreekvaardigheidstoetsing moeten krijgen. Volgens Fulcher vormen intonatiefouten (p. 26) een grotere risicofactor voor spraakverwarring.

Twee begrippen die veelvuldig gebruikt worden in de literatuur over spreekvaardigheid zijn *accuracy* (nauwkeurigheid) en *fluency* (vloeiendheid). Een hoge mate van nauwkeurigheid houdt bijvoorbeeld in dat een spreker grammaticaal correct spreekt, een goede uitspraak hanteert, zijn boodschap duidelijk structureert, en een juiste woordvolgorde gebruikt.

Ook het gebruik van complexe zinnen kan duiden op een hoge mate van nauwkeurigheid en, daarmee samenhangend, een hoger taalniveau, al kan het ook voorkomen dat leerlingen experimenteren met het gebruiken van bijzinnen en daarbij juist meer fouten maken, wat de nauwkeurigheid niet ten goede komt (p. 28). De *syntactic maturity*-hypothese van Hunt (1965) beschrijft het concept van complexere syntactische structuren als graadmeter voor het taalvaardigheidsniveau.

Een hoge mate van vloeiendheid betekent dat iemand op een natuurlijke, ritmische manier spreekt. Minder vloeiendheid in spraak (*disfluency*) uit zich bijvoorbeeld door het gebruik van al dan niet gevulde pauzes, het herhalen van woorden en/of syllabes, en het halverwege een uiting veranderen van grammaticale structuur. Fulcher (p. 30-31) merkt op dat deze fouten er ook op kunnen wijzen dat een leerling een ontwikkeling naar een hoger vaardigheidsniveau doormaakt. Een pauze kan bijvoorbeeld gebruikt worden om na te denken over de zinsstructuur, maar ook om de geplande uiting inhoudelijk te ordenen.

De begrippen *accuracy* en *fluency* worden vaak tegenover elkaar gezet op een continuum; een spreker die een grote mate van nauwkeurigheid laat zien zou dan minder vloeiend kunnen gaan spreken, en meer vloeiende spraak kan ten koste gaan van de nauwkeurigheid van de uiting.

Ook De Jong et al. (2012) gaan uit van een componentiële structuur van spreekvaardigheid. De auteurs toetsten spreekvaardigheid aan de hand van 8 deelvaardigheden, namelijk: kennis van vocabulaire, kennis van grammatica, de snelheid waarmee lexicale informatie wordt opgehaald, de snelheid van articulatie, de snelheid waarmee zinnen opgebouwd worden, uitspraak, klemtoon, en intonatie. Aan dit onderzoek deden 181 T2-leerders van het Nederlands mee. Er was een controlegroep van 54 moedertaalsprekers van het Nederlands (p. 22). De auteurs vonden dat kennis

van vocabulaire en het gebruik van correcte intonatie de beste voorspellers van spreekvaardigheid waren. Dit betekent natuurlijk niet dat de andere factoren niet van belang zijn. De auteurs concluderen dat spreekvaardigheid een geheel is van kennis (van vocabulaire en grammatica), het vermogen om die kennis snel te verwerken (om bijvoorbeeld zinnen te construeren), en uitspraak (van woorden, klemtoon, en intonatie) (p. 29).

Bij de toetsing van spreekvaardigheid is het van belang goed voor ogen te hebben wat het doel van de toets is. Van Gelderen (1992) stelt dat het oplezen van een stuk tekst niets zegt over de manier waarop een spreker zich staande kan houden in een discussie. Luoma (2004) stelt dat een taak waarin ongeplande, spontane spraak centraal staat, zoals het houden van een discussie, een spreker cognitief weliswaar meer belast, maar wel natuurlijker is dan het houden van een monoloog waarbij de spreker van tevoren kan bedenken wat hij of zij wil zeggen.

Van Gelderen (1992) is ook voorstander van toetsvormen die realistische communicatieve situaties zo goed mogelijk simuleren. De auteur noemt in dit verband de zogenoemde geïntegreerde taak, waarin een vaardigheid in samenhang met andere vaardigheden wordt getoetst (al worden die andere vaardigheden daarbij niet beoordeeld). Binnen zo'n taak worden situaties nagebootst die zo herkenbaar mogelijk moeten zijn voor de leerling. Het kan bijvoorbeeld gaan om het opvoeren van een geïmproviseerd toneelstukje. Naarmate zo'n taak de werkelijkheid meer benadert, zou die beter geschikt zijn om taalvaardigheid in communicatieve situaties te meten (p. 6-7).

Van Gelderen noemt dit ecologische validiteit, waarmee hij zich afzet tegen het gebruik van deelvaardigheidstoetsen. Die zouden namelijk geen inzicht bieden in de mate waarin leerlingen zich in maatschappelijke situaties kunnen redden. Logischerwijs brengt deze manier van toetsing ook nadelen met zich mee omdat de deelvaardigheden die wel worden aangewend maar niet beoordeeld, mogelijk invloed kunnen hebben op de vaardigheid die centraal staat in de beoordeling (p. 7).

Het gebruik van een gespreksvorm komt in de literatuur herhaaldelijk naar voren als een van de beste manieren om een goed beeld te krijgen van het niveau van spreekvaardigheid. Rijlaarsdam (1982) noemt als ander voordeel dat bij het voeren van een debat ook de luistervaardigheid getraind wordt. Een ander voordeel van deze vorm is dat het minder spanning oplevert voor de leerlingen, zodat hun spraak vloeiender en natuurlijker is. Ook kunnen meerdere leerlingen tegelijkertijd aan hun spreekvaardigheid werken (Van Gelderen, 1992). De vraag of toetsing met behulp van een computer in plaats van met fysiek aanwezige gesprekspartners afdoende is, is ook van belang, omdat gespreksvaardigheden een wezenlijk onderdeel zijn van de spreekvaardigheid (Iwashita et al., 2008)

Bij het beoordelen van spreekvaardigheid wordt vaak gebruik gemaakt van scoringsrubrieken. Met behulp van deze methode kunnen alle relevante deelvaardigheden afzonderlijk beoordeeld worden op een schaal die verschillende vaardigheidsniveaus aangeeft (bijvoorbeeld een schaal van 1 tot en met 5, waarbij 1 het laagste vaardigheidsniveau representeert, en 5 het hoogste). Welke, en hoeveel, criteria gebruikt worden, is afhankelijk van het doel van de toets. Er kan bijvoorbeeld gekeken worden naar de begrijpelijkheid, uitspraak, grammaticale correctheid, vloeiendheid, en mate van interactie (Nienhuis, 1977), maar de lijst van criteria zal uit praktische overwegingen zelden uitputtend zijn.

Moskal (2000) en Baten et al. (2008) zijn voorstander van het gebruiken van een beoordelingsvorm waarbij wordt omschreven over welke vaardigheden een leerling beschikt, zonder dat daarbij wordt geoordeeld of die leerling een vaardigheid 'goed' of 'fout' doet. Jonsson & Svingby (2007) stellen daarnaast dat het gebruik van beoordelingsrubrieken ervoor zorgt dat het voor leerlingen duidelijker is wat er precies van hen verwacht wordt. Ook dit zou eventuele spanning kunnen verminderen.

Het toetsen van spreekvaardigheid werd lange tijd uitgevoerd aan de hand van interviews tussen twee of meerdere personen. Een van de bekendste formats die hiervoor gebruikt zijn, is het *Main Suite*-systeem. Dit systeem wordt bijvoorbeeld gebruikt bij toetsing voor het Cambridge First English Certificaat (FEC) (O'Sullivan et al. 2002, p33). Dit systeem bestaat uit drie taken die worden uitgevoerd door een interviewer en twee kandidaten. De eerste taak bestaat eruit dat een kandidaat vragen beantwoordt; de tweede taak uit het uitvoeren van een taak door de twee kandidaten op basis van verbale instructies; de derde taak uit een langere discussie tussen de interviewer en de twee kandidaten (p. 34).

Om de validiteit bij dit soort toetsing te waarborgen, moet rekening gehouden worden met een aantal variabelen. Volgens Milanovic & Saville (1996) zijn dat achtereenvolgens de kandidaat, de gesprekspartner, de criteria voor beoordeling, de aard van de taak zelf, en ten slotte de interactie tussen al deze elementen. De validiteit is uiteindelijk afhankelijk van de mate waarin het niveau van de geëliciteerde spraak overeenkomt met het door de onderzoekers beoogde niveau (O'Sullivan et al. 2002, p. 37). Met andere woorden: als er geen gedetailleerde beschrijving voorhanden is van welke soort taalgebruik er getoetst wordt, is het onmogelijk de validiteit van een toets vast te stellen (Anastasi, 1988).

Het analyseren van de mate van overeenkomst tussen verwacht niveau en daadwerkelijk productieniveau van taal kan op een aantal verschillende manieren. Lazaraton (2000) maakt gebruik van *Conversational Analysis*, waarbij een zeer gedetailleerde transcriptie wordt gemaakt van het

taalgebruik in interviews voor spreekvaardigheidstoetsing. Dit is echter een uiterst complex en tijdrovend karwei, dat eigenlijk alleen op kleine schaal werkbaar is, waardoor het ook niet mogelijk is om op basis hiervan te generaliseren.

O'Sullivan et al. (2002, p. 39) zochten naar een manier om een vergelijkbaar resultaat te bereiken met een methode die in real-time toepasbaar was, en stelden hiervoor een checklist samen voor gebruik bij de FCE-toets. (Het niveau van deze toets ligt op ongeveer hetzelfde niveau als niveau B2 van het ERK.²)

Deze checklists zijn nog steeds erg uitgebreid, omdat een valide toets altijd vraagt om bewijs voor taalvaardigheid op meerdere niveaus. Het gaat hier bijvoorbeeld om vaardigheden die variëren van het uiten van een mening, het geven van een beschrijving, en het initiëren van een gesprek (p. 54). Om te zorgen voor een consistente en betrouwbare beoordeling moeten beoordelaars die met deze checklists werken dan ook zorgvuldig getraind worden met behulp van gestandaardiseerde trainingsmaterialen (p. 46).

Meer recent is het gebruik van automatische spraakherkenningssoftware in spraakvaardigheidstoetsing in populariteit gestegen. De Jong et al. (2009) beschrijven de ontwikkeling van de Toets Gesproken Nederlands (TGN). Deze toets wordt gebruikt in het inburgeringsexamen en geeft naast een meting het spraakvaardigheidsniveau ook een beeld van het luistervaardigheidsniveau (p. 43).

De TGN meet taalvaardigheid op ERK-niveaus A1-minus (één niveau onder A1) en A2 (p. 44). Dit niveau zou te vergelijken zijn met dat van een toerist die een bijzondere motivatie heeft om een taal te leren (North, 2000; De Jong et al. 2009, p. 43). De toets kent drie soorten items: zinnen die door de deelnemer herhaald moeten worden, vragen waarop een kort antwoord gegeven moet worden, en woorden waarvan de deelnemer het antoniem moet geven. Er wordt automatisch beoordeeld op twee subsets van criteria: vocabulaire en zinsbouw (inhoud), en uitspraak en vloeiendheid (vorm) (p. 44-45). Volgens Stronks (2011) wordt 75% van het totaalcijfer bepaald door de nazegzinnen.

De onderzoekers vonden bewijs dat de TGN een betrouwbaar middel is om spreekvaardigheid van het Nederlands te meten. Er was geen indicatie dat taal-irrelevante factoren invloed hadden op de uitslag van de toets (De Jong et al. 2009, p. 59).

Kessens & Aarts (2013) deden in opdracht van het Ministerie van Sociale Zaken en Werkgelegenheid namens TNO onderzoek naar de verschillen tussen menselijke en automatische

² Zie het schema op de website <http://www.erk.nl/docent/toetsing/taalcertificaten/Engels/>.

scoring binnen de TGN. Zij vonden geen significante verschillen tussen gemiddelde totaalscores. Wel was het slagingspercentage bij beoordeling via de computer lager dan bij menselijke beoordeling (respectievelijk 51% en 74%). Ook bleek dat de beoordeling consistent was, zowel bij herhaaldelijke meting als bij meting van dezelfde deelnemers (p. 27).

Meer recente ontwikkelingen zullen er in de toekomst voor zorgen dat ook meer complexe communicatieve vaardigheden automatisch getoetst en beoordeeld zullen worden. Litman et al. (2018) geeft een overzicht van de componenten van zogenaamde *spoken dialogue systems*, die voortbouwen op automatische spraakherkenning (p. 6). De auteurs voorspellen dat men in de toekomst, naast intermenselijke communicatieve vaardigheden, ook de communicatie tussen mens en computer wil kunnen beoordelen (p. 12).

3.5 Luistervaardigheid

Luistervaardigheid wordt, uit de vier deelaspecten van taalvaardigheid, misschien wel het minst goed begrepen (Vandergrift & Goh, 2009, p. 22). Er is daarom geen eenduidige definitie van wat luistervaardigheid precies is. In de literatuur zijn verschillende benaderingen te vinden van pogingen om luistervaardigheid te analyseren en te definiëren. Hieronder komen enkele van deze onderzoeken aan bod.

Jones & Plass (2002) verdelen het luisterproces in drie deelprocessen. Allereerst ontvangt een luisteraar input in de vorm van spraak. Die input wordt onderworpen aan zowel linguïstische kennis van de luisteraar als zijn of haar kennis over het gespreksonderwerp. Op basis hiervan wordt betekenis toegekend aan de spraak. De eigen kennis die een luisteraar heeft over het gespreksonderwerp wordt *privileged ground* genoemd. De kennis die beide partijen hebben wordt *common ground* genoemd, en die kennis is nodig voor beide partijen om een betekenisvol gesprek te kunnen voeren. Om de *common ground* te kunnen vergroten is interactie nodig tussen de gespreksdeelnemers.

Interactie wordt algemeen beschouwd als een essentieel onderdeel van interpersoonlijke communicatie, en wordt dan ook van groot belang geacht in het taalleerproces. Baten et al. (2008, p. 9) geven aan dat het ontwikkelen van een algemene communicatieve competentie (in een tweede taal) binnen het Europees Referentiekader van groter belang wordt geacht dan het ontwikkelen van deelaspecten van taalvaardigheid.

Volgens Varonis & Gass (1985, p. 67) fungeert interactie tussen moedertaalspreker en T2-luisteraars als een manier om de *common ground* te vergroten. Hierbij past de moedertaalspreker

zijn of haar niveau aan naar het niveau van de T2-luisteraar. Anderzijds kan een T2-luisteraar ook bepaalde strategieën toepassen om hetzelfde doel te bereiken, zoals het herhalen van de spraak van de gesprekspartner. Murphey (2001) noemt dit *conversational shadowing*. Er kan sprake zijn van het herhalen van de complete uiting (*complete shadowing*), van een deel van de uiting (*selective shadowing*, waarbij bijvoorbeeld alleen de laatste woorden van een zin worden herhaald), en van *interactive shadowing*, waarbij de luisteraar woorden herhaalt, en daar bijvoorbeeld vragen en eigen commentaar aan toevoegt (p. 129). *Complete shadowing* zal vaker voorkomen bij leerlingen met een lager vaardigheidsniveau, terwijl *interactive shadowing* alleen voorkomt op hogere vaardigheidsniveaus. Dergelijke woordherhaling is onderdeel van natuurlijke spraak, en geeft aan dat een luisteraar de uiting van een gesprekspartner verstaat en begrijpt, zodat beide gesprekspartners weten dat ze zich op een zelfde niveau van gespreksvaardigheid bevinden (Tannen 1989, p. 13).

Rost (2002) benadert luistervaardigheid vanuit zowel een cognitieve als een sociale dimensie. Bezien vanuit de cognitieve dimensie bestaat luistervaardigheid uit een drietal fundamentele principes. Ten eerste moet de luisteraar de aandacht vestigen op de gehoorde spraakstroom, om in deze input woorden te herkennen en die vervolgens te parseren tot betekenisvolle eenheden. Voor de verwerking van de op deze manier verkregen informatie wordt bestaande kennis uit het langetermijngeheugen aangesproken. Op deze manier kan gebrekkig begrip bijvoorbeeld overbrugd worden door gebruik te maken van contextuele informatie. Ten slotte stelt Rost dat de snelheid van de verwerking van gesproken tekst afhankelijk is van het totale taalvaardigheidsniveau. Leerlingen met een hoger taalvaardigheidsniveau zullen spraak dus sneller verwerken (p. 2).

Binnen de sociale dimensie van luistervaardigheid stelt Rost dat ook andere input van belang is, zoals lichaamstaal (p. 5) en verschillen in status tussen gesprekspartners (p. 6). Ook waarschuwt de auteur voor een al te grote focus op taalproductie tijdens lessen en toetsing. Dit zou ervoor kunnen zorgen dat een leerling last krijgt van *listening anxiety*, wat een negatieve invloed kan hebben op luistervaardigheidstoetsing. Op zijn beurt zou dit kunnen leiden tot een lagere mate van motivatie, waardoor een leerling gebruik gaat maken van minder effectieve luisterstrategieën (p. 7).

Luistervaardigheid wordt op verschillende manieren getoetst. Uit de literatuur blijkt dat communicatieve interactie van grote invloed is op de luistervaardigheid. Een communicatieve benadering in de toetsing lijkt dan ook voor de hand te liggen om een zo goed mogelijk beeld te krijgen van de luistervaardigheid van een leerling.

Logischerwijs is het niet altijd mogelijk om luistervaardigheid te toetsen aan de hand van een echt gesprek. In veel gevallen wordt gebruik gemaakt van de computer om een

luistervaardigheidstoets af te nemen. Dit is wellicht gemakkelijker wanneer er veel leerlingen getoetst moeten worden, maar het heeft ook een aantal nadelen.

Krashen (1981) deed onderzoek naar het gebruik van videobeelden bij de toetsing van luistervaardigheid. De auteur hanteert de *input hypothesis*, die stelt dat een leerling authentieke input nodig heeft van twee niveaus: één deel van de input bevindt zich op het taalniveau dat de leerling aankan, een overig deel bestaat uit spraak van een niveau dat nog net buiten zijn of haar bereik ligt. De videobeelden worden gebruikt om een zo authentiek mogelijk gesprek te simuleren. Er kan bijvoorbeeld sprake zijn van 'oogcontact', en de leerling ziet de lichaamstaal van de spreker.

Volgens Richardson & Scinincariello (1989) biedt deze visuele input de leerling de nodige contextuele houvast om ook de input die net buiten het vaardigheidsniveau ligt te begrijpen. Daarnaast stellen deze auteurs dat het gebruik van visuele hulpmiddelen een actieve leerhouding stimuleert. Wel adviseren zij de lengte en inhoud van de gebruikte fragmenten af te stemmen op het niveau van de individuele leerling, en ervoor te zorgen dat de visuele input de gesproken tekst ondersteunt.

Recent onderzoek van Den Boer et al. (2011) toonde aan dat gebruik van videobeelden bij luistertoetsen zorgde voor een significant hogere motivatie van de leerlingen. Het zo nauwkeurig mogelijk simuleren van een communicatieve situatie zorgde er ook voor dat leerlingen hoger scoorden.

In de literatuur is ook beperkt aandacht voor zelfbeoordeling van de luistervaardigheid. Boers (2008) deed kleinschalig onderzoek naar de mate waarin NT2-cursisten hun eigen luistervaardigheidsniveau inschatten. Dit werd afgezet tegen de oordelen van drie NT2-docenten. Uit eerder onderzoek (Janssen-van Dieten, 1992; Van Onna en Jansen, 2006) bleek dat zowel leerders van Nederlands als tweede taal als Nederlandse moedertaalsprekers die een andere taal leren, hun eigen taalniveau systematisch te hoog inschatten.

Een groep van 21 hoogopgeleide volwassen beginnende leerders van het Nederlands als tweede taal werd na 9 weken les gevraagd hun eigen lees- en luistervaardigheid in te schatten. Drie NT2-docenten deden hetzelfde voor deze leerlingen. Hierna werd bij de leerlingen een lees- en luistertoets afgenomen.

Na analyse bleek dat meer dan de helft van de leerlingen hun lees- en luistervaardigheid te hoog inschatte. De docenten schatten bij minimaal 8 leerlingen de lees- en luistervaardigheid te hoog in. Boers concludeert dan ook dat de neiging tot (zelf)overschatting bij het gebruiken van

zelfbeoordeling aanwezig is. Hiermee moet rekening worden gehouden wanneer er gebruik wordt gemaakt van zelfevaluatie, bijvoorbeeld bij het bijhouden van een portfolio.

Op het gebied van luistervaardigheidstoetsing is relatief minder specifieke literatuur te vinden. Volgens Bonk (2000, p. 15) komt dit doordat luistervaardigheidsonderzoek vooral leunt op bevindingen uit de literatuur over leesvaardigheidsonderzoek.

Wel lijkt het erop dat deze situatie de werkelijkheid weerspiegelt. Leesvaardigheid en luistervaardigheid ondersteunen elkaar. Bij T1-gebruikers is dit waarschijnlijk meer het geval dan bij T2-leerders. De laatste groep vertoont betere resultaten wanneer de vraagstelling geschreven is in plaats van gesproken (Lund, 1991), en verstaat vaak zelfs de hele gesproken tekst niet (Hirai, 1999).

Een mogelijke verklaring hiervoor is een gebrek aan voldoende groot vocabulaire, en een daarmee samengaande onvoldoende vaardigheid in woordherkenning (Bonk 2000, p. 16). Aan de hand van de studie van Bonk is niet duidelijk vast te stellen hoe groot iemands lexicale kennis moet zijn om een luistertoets met succes af te leggen. Wel is er een correlatie aan te wijzen tussen een hoge mate van kennis en een verhoogde kans op slagen, al is dit verband ook niet absoluut (p. 24). Bonk merkt op (p. 19; 25) dat in de studie gebruik werd gemaakt van opnames van voorgelezen tekst, en niet van natuurlijke spraak, wat de resultaten wellicht heeft beïnvloed.

Schmidt-Rinehart (1994) vond, naast een significante relatie tussen luistervaardigheid en T2-trainingsniveau, dat achtergrondkennis ook een rol speelt bij luistervaardigheid. De auteur vond een significant verschil in resultaten tussen deelnemers met verschillende niveaus van achtergrondkennis over het onderwerp waarover zij een tekst te horen kregen. Hoe meer achtergrondkennis, des te beter waren de behaalde resultaten. Dit effect was zichtbaar bij studenten van alle onderzochte vaardigheidsniveaus. De auteur waarschuwt er wel voor dat de concepten taalvaardigheid en taalbegrip niet met elkaar verward of gelijkgesteld moeten worden (p. 185).

Wanneer een luistertoetsontwerper deelnemers een authentieke afspiegeling van de T2 wil voorzetten, krijgt men onvermijdelijk te maken met regionale spraakkenmerken. Major et al. (2002) onderzochten de invloed van accenten op taalbegrip bij luistertoetsen. Volgens de auteurs wordt algemeen aangenomen dat accenten die het eigen accent dichter benaderen, beter worden begrepen (p. 174). Hiernaast speelt de attitude van de deelnemer ten opzichte van de gehoorde accenten ook een rol (p. 175), en wordt aangenomen dat grotere bekendheid met buitenlandse accenten helpt bij het ontwikkelen van begrip (Tauroza & Luk, 1997).

In de literatuur wordt verder opgemerkt dat, hoewel dit onderzoeksgebied minder aandacht krijgt dan de overige gebieden van taalvaardigheid, het wel degelijk meer aandacht verdient. McLean

et al. (2015) signaleerden een lacune in de literatuur wat betreft vocabulaireluistertoetsen. Zij menen dat deze toetsen wel degelijk van belang zijn, en wel om drie redenen. Ten eerste kan de orthografische en fonologische lexicale kennis van een T2-leerder erg uiteenlopen. Ten tweede doet een online luistertoets een groter beroep op het werkgeheugen dan een offline schrijftoets. Ten derde spreekt een luistertoets ook andere elementen aan dan een schrijftoets, zoals kennis van fonologie en prosodie van de doeltaal.

4 Inburgering in de praktijk

In dit onderzoek staat de taaltoetsing in het inburgeringsproces centraal. In dit hoofdstuk wordt eerst de bredere context van inburgering behandeld. Daarna worden de taaltoetsen in het traject besproken.

In sectie 4.1 wordt een korte beschrijving gegeven van de inburgeringsprocedure in Nederland en de eisen waaraan voldaan moet worden voordat een verblijfsvergunning verleend kan worden. In sectie 4.2 wordt de rol die taal speelt in het inburgeringsproces behandeld. In het kader van de internationale context wordt in sectie 4.3 een globaal beeld gegeven van de inburgeringspraktijk in een select aantal andere landen en van de overeenkomsten en verschillen met die van Nederland. Ook komen internationale afspraken op het gebied van inburgering aan bod.

In sectie 4.4 wordt ingegaan op de taaltoetsen die in de inburgeringsprocedure in Nederland gebruikt worden. Hierbij staat de manier waarop de verschillende soorten taalvaardigheid precies getoetst worden en de taalkundige inzichten die de fundering van het toetsontwerp vormen centraal.

4.1 De inburgeringsprocedure in Nederland

Een groot deel van de immigranten die vanuit het buitenland naar Nederland komen, moet inburgeren. Dit is een voorwaarde voor het verkrijgen van een verblijfsvergunning. Mensen die voor langere tijd in Nederland willen verblijven zijn daarom verplicht een inburgeringsexamen af te leggen.

Dit examen is niet voor iedere immigrant verplicht. Op de website van de Rijksoverheid³ is te lezen voor welke vreemdelingen het inburgeringsexamen precies bedoeld is. Onder meer Immigranten uit de Europese Unie (EU), de Europese Economische Ruimte (EER)⁴, Turkije⁵ en Zwitserland zijn bijvoorbeeld niet verplicht in te burgeren, evenmin als immigranten jonger dan 18 jaar of ouder dan de pensioengerechtigde leeftijd. Ook mensen die over Nederlandse diploma's beschikken zijn uitgezonderd van deze plicht.

³ <https://www.rijksoverheid.nl/onderwerpen/nieuw-in-nederland/vraag-en-antwoord/moet-ik-als-nieuwkomer-inburgeren>

⁴ De Europese Economische Ruimte bestaat uit de 28 lidstaten van de Europese Unie plus Noorwegen, IJsland, en Liechtenstein. Deze drie landen zijn geen lid van de Europese Unie maar wel van de Europese Vrijhandelsassociatie. (Zwitserland, ook een lidstaat van de Europese Vrijhandelsassociatie, is na een afwijzend referendum niet toegetreden tot de Europese Economische Ruimte.)

⁵ Een associatieovereenkomst tussen Turkije en de Europese Economische Gemeenschap uit 1963 bepaalt dat mensen uit Turkije die zich in de Europese Unie willen vestigen, niet verplicht zijn om in te burgeren. De tekst van het verdrag is te vinden op deze website: <http://wetten.overheid.nl/BWBV0004316/2004-05-01>

Iedere immigrant voor wie inburgering wel verplicht is, wordt geacht al in het land van herkomst een basisexamen inburgering af te leggen. Dit geldt dus ook voor mensen die in het kader van gezinshereniging naar Nederland willen komen.

Inburgering in Nederland behelst drie plichten waaraan een immigrant moet voldoen. Allereerst moet hij de Nederlandse taal leren. Ten tweede moet hij leren hoe de Nederlandse samenleving werkt en hoe men in Nederland met elkaar omgaat. De derde voorwaarde is actieve participatie: de immigrant wordt bijvoorbeeld geacht te werken of een opleiding te volgen.

In het geval van migranten die niet als asielzoeker naar Nederland komen, geldt de plicht al in het land van herkomst het *basisexamen inburgering in het buitenland* af te leggen. In dit examen wordt getoetst op spreek- en leesvaardigheid van het Nederlands en kennis van de Nederlandse samenleving. Dit examen wordt afgenomen met behulp van een computer. Voor het spreekvaardigheids gedeelte worden de antwoorden ingesproken, de overige onderdelen zijn meerkeuzevragen. Het examen duurt ongeveer twee uur. Het niveau waarop getoetst wordt is niveau A1 van het Europees Referentiekader. Dit wordt in sectie 4.2.1 verder besproken.

Zijn alle onderdelen van het basisexamen voldoende afgesloten, dan wordt er een machtiging tot voorlopig verblijf (MVV) afgegeven, waarmee men zich voorlopig in Nederland kan vestigen. Hierna wordt het inburgeringsproces vervolgd met het *inburgeringsexamen in Nederland*. Dit examen heeft een hogere moeilijkheidsgraad dan het basisexamen. Er wordt hier naast spreek- en leesvaardigheid ook getoetst op luister- en schrijfvaardigheid van het Nederlands, en naast een uitgebreider toetsingsonderdeel over kennis van de Nederlandse samenleving bevat het examen ook een onderdeel oriëntatie op de Nederlandse arbeidsmarkt.

Om kennis op te doen die vereist is voor het succesvol afleggen van het inburgeringsexamen kan een speciale inburgeringscursus gevolgd worden. Deze cursussen worden aangeboden door verschillende instanties, maar niet door de overheid zelf. Tijdens zo'n cursus krijgt de immigrant les in de Nederlandse taal en over de Nederlandse maatschappij. Er geldt een termijn van 3 jaar om in te burgeren. De dienst uitvoering onderwijs (DUO) zorgt ervoor dat elke inburgeringskandidaat een brief krijgt met daarin de startdatum van de inburgeringstermijn. Ook biedt DUO de mogelijkheid geld te lenen om taal- en inburgeringscursussen en de deelname aan het inburgeringsexamen te bekostigen.⁶

⁶ Op de website <https://www.inburgeren.nl/inburgeren-betalen/index.jsp> is een gedetailleerd overzicht van de kosten van elk inburgeringsexamen te vinden. De kosten van een inburgeringscursus kunnen variëren, bijvoorbeeld afhankelijk van het aantal cursussen die iemand nodig heeft.

4.1.1 De geschiedenis van inburgering in Nederland

Gysen et al. (2009) geeft een kort overzicht van de manier waarop het inburgeringproces in Nederland is ontstaan. Volgens de auteurs heeft het proces zich ontwikkeld van vrijwel onbestaand naar een vrij strenge procedure, die zou moeten leiden tot een stabiele instroom van immigranten in de komende decennia (p. 99).

Tot bijna aan het eind van de jaren '80 van de vorige eeuw was er nog geen officieel inburgeringsbeleid in Nederland. Op taalgebied werden cursussen verzorgd door vrijwilligers. In 1989 werd de eerste stap gezet tot de ontwikkeling van een taalbeleid vanuit de overheid. Gedurende de jaren '90 nam het gebruik van taaltoetsen toe (p. 98).

In 1998 werd een verplichte inburgeringstoets ingevoerd. Hier waren voor de inburgeraar nog geen kosten aan verbonden. In 2003 werd het voltooiën van een cursus Nederlands op A2-niveau van het Gemeenschappelijk Europees Referentiekader (dat in de volgende sectie wordt besproken) verplicht gesteld voor inburgering. Als iemand een hogere taalvaardigheid kon aantonen, dan hoefde deze toets niet gedaan te worden (p. 100).

In 2006 werd vervolgens een wet aangenomen die het verplicht stelde voor mensen in de leeftijd van 16 tot en met 65 jaar en afkomstig van buiten de EU die zich in Nederland willen vestigen al voor hun aankomst een toets af te leggen. Het ging daarbij enkel om spreek- en luistervaardigheid op het zogenaamde A1-minus-niveau (p. 100).

In een van de volgende secties worden de huidige vormen van het Nederlandse inburgeringsexamen besproken. Opvallend hieraan, in vergelijking met de eerdere wetgeving, is dat het niveau van het basisexamen inburgering is verhoogd van niveau A1-minus (zie sectie 4.2.1) naar niveau A1. Voor het voltooiën van het inburgeringsexamen dat in Nederland afgelegd kan worden, wordt een termijn van 3 jaar gesteld. Aan het overschrijden van die termijn zijn relatief hoge financiële sancties verbonden, namelijk een boete die minstens €250 bedraagt en kan oplopen tot €1250, afhankelijk van onder meer het aantal gehaalde examenonderdelen. Dit proces is voor elke immigrant gelijk.

4.2 De rol van taal in de inburgeringsprocedure: taaltoetsing en taalanalyse

4.2.1 De rol van taal in toetsing en de getoetste taalvaardigheidsniveaus

Het leren spreken, luisteren, lezen, en schrijven van de Nederlandse taal is verreweg het belangrijkste onderdeel van het inburgeringsproces. Toetsing op deze onderdelen maakt dan ook het overgrote deel van het inburgeringsexamen uit.

Om succesvol in te burgeren, wordt getoetst op twee verschillende taalvaardigheidsniveaus. Deze niveaus zijn gedetailleerd vastgelegd in het *Common Framework of Reference for Languages* (CEF), in 2001 gepubliceerd door de Raad van Europa. Dit referentiekader is later door de Nederlandse Taalunie⁷ in het Nederlands gepubliceerd onder de titel *Gemeenschappelijk referentiekader voor moderne vreemde talen: Leren, onderwijzen, beoordelen*.

Het opstellen van een Europees Referentiekader (ERK) had onder meer als doel een gemeenschappelijke norm te formuleren voor het classificeren van taalvaardigheid in het (vreemdetalen)onderwijs. Dit kader kan zo als basis dienen voor het ontwikkelen van onderwijsprogramma's in heel Europa (Baten et al. 2008, p. 7).

Van de zes taalvaardigheidsniveaus (A1, A2, B1, B2, C1, en C2) die in het ERK beschreven staan, zijn alleen A1 en A2 voor dit onderzoek relevant. Niveau A duidt een beginnend spreker aan, een zogenoemde basisgebruiker. Het basisexamen inburgering in het buitenland is dan ook afgestemd op niveau A1. Baten et al. (2008) geven de volgende beschrijvingen van de competenties die een spreker op dit niveau beheerst:

- **Lezen:** Kan zeer korte, eenvoudige teksten frase voor frase begrijpen door vertrouwde namen, woorden, en elementaire combinaties te herkennen en indien nodig te herlezen.
- **Schrijven:** Kan eenvoudige op zichzelf staande frasen en zinnen schrijven.
- **Spreken:** Kan simpele, voornamelijk op zichzelf staande frasen produceren over mensen en plaatsen.
- **Luisteren:** Kan spraak volgen die heel langzaam en zorgvuldig wordt uitgesproken, met lange pauzes om de betekenis op te nemen.

De Jong et al. (2009) geven een definitie van het zogenaamde A1-minus-niveau, dat in opdracht van het Ministerie van Justitie geformuleerd is voor de Toets Gesproken Nederlands, een

^{7 7} De website www.taalunieversum.org/inhoud/erk-nederlands biedt een heldere introductie tot het ERK. Op deze site is ook de Nederlandse vertaling van de oorspronkelijke publicatie van de Raad van Europa te downloaden.

voorloper van het basisexamen inburgering. Dit niveau beschreef een zodanige taalvaardigheid dat een spreker zich met geïsoleerde woorden enigszins verstaanbaar kan maken, maar nog wel een sterk accent heeft. Een spreker op dit niveau begrijpt simpele termen, maar een echte conversatie voeren is niet goed mogelijk. Om een idee te krijgen van het verschil tussen A1-minus en A1 halen De Jong et al. North (2000) aan, die een met A1-minus vergelijkbaar taalvaardigheidsniveau het label 'Tourist' gaf. Simpel gezegd bevindt A1-niveau zich dus vlak boven het taalvaardigheidsniveau van een gemiddelde toerist die de taal niet machtig is.

Het inburgeringsexamen in Nederland, waarin ook luister- en schrijfvaardigheid zijn meegenomen, toetst op taalvaardigheidsniveau A2. Dit niveau wordt door Baten et al. (2008) voor de vier deelvaardigheden als volgt beschreven:

- **Lezen:** Kan korte, eenvoudige teksten begrijpen over vertrouwde zaken van concrete aard, die zijn geschreven in veelgebruikte alledaagse of werkgebonden taal. Kan korte, eenvoudige teksten begrijpen, geschreven met de meest gebruikte woorden, met inbegrip van een aantal woorden uit de gemeenschappelijke internationale woordenschat.
- **Schrijven:** Kan een reeks eenvoudige frasen en zinnen schrijven, verbonden door simpele voegwoorden als 'en', 'maar' en 'omdat'.
- **Spreeken:** Kan een eenvoudige beschrijving of presentatie geven van mensen, woon- of werkomstandigheden, dagelijkse routines, voorkeuren en afkeren en dergelijke in een korte reeks simpele frasen en zinnen die als in een lijst met elkaar verbonden zijn.
- **Luisteren:** Kan genoeg verstaan om te kunnen voldoen aan behoeften van concrete aard, mits er helder en langzaam wordt gearticuleerd. Kan frasen en uitdrukkingen verstaan die verband houden met zaken van de meest directe prioriteit (elementaire persoons- en familiegegevens, boodschappen doen, plaatselijke geografie, werk), mits er helder en langzaam wordt gearticuleerd.

4.2.2 LADO: taalanalyse in de asielprocedure

Language Analysis for the Determination of Origin, oftewel LADO, is een procedure waarin het taalgebruik van vluchtelingen wordt geanalyseerd om te achterhalen of iemand wel echt afkomstig is uit het land of de regio die hij of zij als zodanig claimt.

De LADO-procedure wordt niet altijd gebruikt, maar is een optioneel instrument om bij twijfel meer informatie over de afkomst van een asielzoeker te verkrijgen. Tijdens de procedure wordt een interview afgenomen met de asielzoeker in kwestie, waarin hij of zij wordt verzocht alle talen en taalvarianten die hij of zij beheerst te laten horen. Ook mag de asielzoeker spreken over alles wat mogelijk kan helpen bij het verkrijgen van een verblijfsvergunning.

Van dit gesprek wordt een opname gemaakt die vervolgens wordt gebruikt voor analyse. Een onafhankelijke analist (met specifieke kennis van de taal of talen in kwestie) kijkt naar de taalkenmerken die de asielzoeker laat horen, en doet aan de hand van die kenmerken een uitspraak over de mate waarin aannemelijk is dat de geclaimde regio van afkomst overeenkomstig is met de werkelijkheid. Er wordt nog steeds gediscussieerd over de vraag of een interview met een asielzoeker een goede manier is om de voor analyse nodige informatie te verkrijgen. Over de manier waarop een deskundige zich moet uitspreken over de uitkomst van een analyse is wel overeenstemming (Broeders 2010, p. 51).

Omdat LADO een waardevol taalkundig instrument is in de Nederlandse immigratiepraktijk, wordt het hier kort vermeld. Omdat het echter specifiek van belang is voor de asielprocedure, en niet voor de inburgeringsprocedure, valt het buiten het onderwerp van dit onderzoek en zal het hier niet verder behandeld worden.

4.3 De internationale context: de inburgeringsprocedure in andere landen

4.3.1 Hoe gaan andere landen om met inburgering?

Niet elk land heeft een inburgeringsprocedure die uit zo'n uitgebreid traject bestaat als Nederland. Piller (2001) geeft een overzicht van de inburgeringsprocedure in enkele andere landen.

In Australië zijn alleen basisvaardigheden van het Engels vereist, zodat een immigrant kan begrijpen wat er in de wet staat en weet dat hij of zij stemrecht heeft, in een jury kan optreden, en zo nodig in militaire dienst kan. Het belangrijkste criterium om de mate van taalvaardigheid vast te stellen, is het met goed gevolg doorstaan van een interview met een medewerker van het departement van immigratie. Wanneer het interview zonder hulp van een tolk afgenomen kan worden, wordt de taalvaardigheid van de immigrant voldoende geacht (p. 266).

De inburgeringsprocedure in de Verenigde Staten lijkt op die van Nederland. De vier domeinen van taalvaardigheid worden getoetst: immigranten lezen hardop een tekst, schrijven enkele zinnen, en beantwoorden mondeling vragen over zichzelf en hun verblijfsaanvraag (p. 266). Daarnaast is een toetsingsonderdeel met vragen over de Amerikaanse maatschappij een verplicht onderdeel van de procedure.⁸

⁸ <https://www.uscis.gov/us-citizenship/naturalization-test>

In Frankrijk zijn burgerschap en nationaliteit onlosmakelijk met elkaar verbonden, en de taal is daar een integraal onderdeel van de nationale identiteit. Zonder Frans te spreken en zich de Franse cultuur eigen te maken kan iemand niet de Franse nationaliteit hebben. Het goed spreken van de Franse taal wordt dan ook als vereiste gezien om staatsburger te kunnen worden; aan Franstaligen uit andere landen worden dan ook erg soepele inburgeringseisen gesteld. De situatie in Duitsland is vergelijkbaar, al ligt de nadruk in dat land meer op afkomst dan op taalvaardigheid (p. 268). Bewijs van Hebreeuwse taalvaardigheid is niet vereist voor immigranten die in Israël willen verblijven, maar religie en etniciteit is des te belangrijker (p. 265). Zo worden in de verschillen in taalbeleid in de immigratiewetgeving tussen landen onderling duidelijk bepaalde ideologieën van nationaliteit uitgedragen.

4.3.2 Verschillen tussen de inburgeringsprocedures van Nederland en Vlaanderen

Hoewel in hetzelfde taalgebied, zijn er opvallende verschillen aan te wijzen tussen het inburgeringsbeleid van Nederland en dat van Vlaanderen. Hieronder volgt allereerst een korte beschrijving van het verloop van het inburgeringsproces in Vlaanderen.

Op het eerste gezicht lijkt de inburgering in Vlaanderen erg op die in Nederland. Het inburgeringsprogramma⁹ bestaat uit vier onderdelen, waarvan twee cursussen: een taalcursus Nederlands, en een cursus waarin de inburgeraar kennismaakt met het leven in en de maatschappij van België en Vlaanderen. De laatstgenoemde cursus wordt aangeboden in de moedertaal van de immigrant, of in een taal die hij of zij machtig is.

Daarnaast krijgt elke inburgeraar een individuele begeleider toegewezen die helpt bij het zoeken naar werk of een opleiding. Ook krijgen zij informatie over sport, cultuur, en vrije tijdsbesteding in hun nieuwe land van vestiging. De begeleiders spreken altijd de taal van de inburgeraar, of een taal die hij of zij begrijpt. Deze individuele begeleiding vindt echter pas plaats nadat de twee cursussen met goed gevolg zijn afgelegd en een *attest van inburgering* is behaald.

De inburgeringsprocedure is bedoeld voor nieuwe immigranten van 18 jaar of ouder – minderjarige immigranten burgeren in via het reguliere onderwijs – die voor langere tijd in Brussel of Vlaanderen willen komen wonen. Immigranten van wie één van de ouders Belg is, komen ook in aanmerking voor deelname aan de procedure, al is het voor hen niet altijd verlicht in te burgeren. In tegenstelling tot het inburgeringsproces in Nederland, waar de kosten voor rekening van de immigrant zijn, is het traject in België gratis.

⁹ Een gedetailleerd overzicht van het Vlaamse inburgeringstraject is te vinden op deze website: <https://www.vlaanderen.be/nl/gemeenten-en-provincies/dienstverlening-van-gemeenten-en-provincies/begeleiding-van-inburgeraars-het-inburgeringstraject>

De Belgische overheid biedt zelf geen taalcursussen aan, maar geeft wel de mogelijkheid om de best passende school en cursus te vinden.¹⁰ Op de website integratie-inburgering.be wordt een oefentaaltoets Nederlands aangeboden.¹¹

Met ingang van 16 augustus 2017 wordt gebruik gemaakt van nieuwe taaltoetsen, die worden afgenomen tijdens twee toetsmomenten. Tijdens het eerste toetsmoment staat de receptieve taalvaardigheid (luisteren, lezen) centraal, tijdens het tweede de productieve taalvaardigheid (spreken, schrijven). Men moet slagen voor het eerste toetsmoment alvorens door te kunnen stromen naar het tweede. Nadat beide onderdelen met goed gevolg afgelegd zijn, krijgt de inburgeraar een officieel attest.

Anders dan in de Nederlandse inburgeringsprocedure wordt er niet alleen getoetst op de taalvaardigheidsniveaus A1 en A2, maar ook op de niveaus B1 en B2. Inburgeraars kunnen zelf kiezen op welk niveau zij getoetst willen worden. Die keuze is afhankelijk van de reden waarvoor zij hun attest willen gebruiken: om alleen te inburgeren is A1 vereist, voor een nationaliteitsaanvraag niveau A2. Wil een immigrant bijvoorbeeld gaan werken in het volwassenonderwijs, dan is niveau B1 of B2 (of hoger) vereist.

4.4 Toetsing in de Nederlandse inburgeringsprocedure

4.4.1 De in de Nederlandse inburgeringsprocedure gebruikte toetsen

Immigranten die op of na 1 oktober 2017 beginnen met inburgeren, zijn verplicht een participatieverklaring te behalen voordat zij kunnen deelnemen aan de inburgeringsexamens. De participatieverklaring is een document waarin de immigrant verklaart de in Nederland geldende waarden, normen, rechten, en plichten te onderschrijven, en actief mee te zullen doen aan de Nederlandse samenleving. Na inschrijving bij een gemeente geldt een periode van 1 jaar om te leren wat deze verklaring inhoudt en die te ondertekenen. Gebeurt dit niet, dan krijgt de immigrant een boete van DUO, en kan er ook geen geld meer geleend worden van DUO.¹²

¹⁰ <https://www.integratie-inburgering.be/nederlands-leren>

¹¹ <https://www.integratie-inburgering.be/taaltest-nederlands>

¹² Meer informatie over de inhoud van de participatieverklaring is te vinden op de website <https://www.inburgeren.nl/participatieverklaring.jsp>

Tenzij de immigrant al eerder een NT2-certificaat heeft behaald, zijn er vier taalvaardigheidsexamens die behaald moeten worden. Er wordt getoetst op lees-, schrijf-, spreek-, en luistervaardigheid.

Ten slotte moet er een examen *Kennis van de Nederlandse Samenleving* afgelegd worden, en moet er een portfolio met betrekking op *Oriëntatie op de Nederlandse Arbeidsmarkt* aangelegd worden.¹³ Omdat dit onderzoek zich beperkt tot de taalvaardigheidstoetsen in de inburgering, zullen de twee laatstgenoemde examens hier niet verder worden behandeld.

4.4.2 Examenonderdeel leesvaardigheid

Het examenonderdeel leesvaardigheid wordt afgenomen met behulp van een computer. Voor het voltooien van dit onderdeel heeft elke deelnemer 50 minuten de tijd. Het examen bestaat uit het lezen van korte teksten en het beantwoorden van ongeveer 25 daarbij horende vragen.

Hieronder (voorbeeld 4.1) volgt een van de opdrachten uit het oefenexamen. Het bestaat uit een tekst (casus) met een bijbehorende vraag en een aantal mogelijk antwoorden.¹⁴

4.1 Casus

Michael werkt op een kantoor.
Zijn collega Nadia gaat trouwen met Richard.
Michael krijgt een e-mail van de feestcommissie.

Lees eerst de vraag.
Lees daarna de tekst.

Vraag

Michael wil met de bus mee naar de bruiloft.
Bij wie moet hij zich opgeven?

Tekst

Bruiloft Nadia en Richard

Beste collega's,

Volgende week vrijdag is de bruiloft van Nadia en Richard! Daarom hebben we een

¹³ Meer informatie over het examenonderdeel *Oriëntatie op de Nederlandse Arbeidsmarkt* is te vinden op de website <https://www.inburgeren.nl/examen-doen/examen-ona.jsp>

¹⁴ <http://www.oefenexamensduo.nl/portal/appmodules/examen/startexamen.ctrl>

feestcommissie opgericht. De feestcommissie bestaat uit William, David en Elise. We hebben alle drie een taak, zodat we er een leuke dag van kunnen maken!

Wij willen een bus huren om samen met collega's naar de bruiloft te gaan. Wil je mee met de bus? Geef je dan op bij David.

We willen met de collega's een cadeau geven aan het bruidspaar. Als je wilt, kun je ook meedoen. Je kunt geld aan William geven. We weten nog niet precies wat we gaan geven. Op het bureau van Elise ligt een lijstje met cadeaus. Kruis op het lijstje aan wat je het leukste cadeau vindt: een cadeaubon of een weekendje weg. We gaan het cadeau geven dat het meest gekozen is door jullie.

Nadia is deze week nog gewoon aan het werk. Ons cadeau moet natuurlijk wel een verrassing blijven. Let er dus op dat je niets zegt tegen Nadia!

Groeten, de feestcommissie
David, William en Elise

Antwoord

A: bij William

B: bij Elise

C: bij David

Andere voorbeelden van teksten zijn cursusbeschrijvingen, lijsten met tips en teksten uit folders. Deze teksten zijn, gezien de lengte en de complexiteit, afgestemd op het gewenste A2-niveau. Helaas is voor dit oefenexamen geen nakijkmodel beschikbaar.

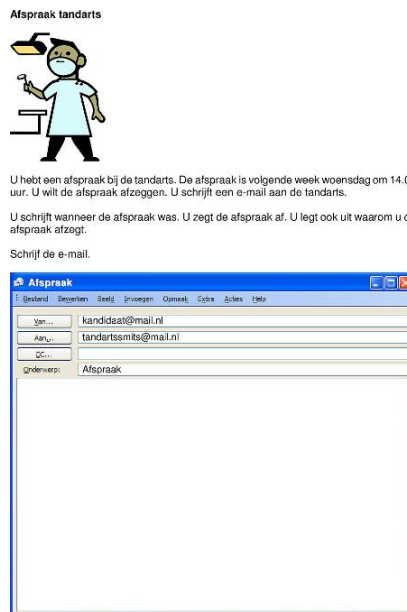
4.4.3 Examenonderdeel schrijfvaardigheid

Het examenonderdeel schrijfvaardigheid wordt, anders dan de andere examenonderdelen van het inburgeringsexamen, niet afgelegd met behulp van een computer, maar met pen en papier.

Deelnemers krijgen 35 minuten de tijd om een aantal opdrachten te maken. Daarbij horen onder meer het afmaken van zinnen, zoals de zin voorbeeld (1)¹⁵:

1 Paula is op school. De lessen zijn afgelopen, daarom _____.

Naast afmaakoefeningen bevat de toets opdrachten waarbij een brief of email geschreven moet worden (figuur 4.1) en een formulier ingevuld moet worden (figuur 4.2).



Figuur 4.1: oefenopgave examenonderdeel schrijfvaardigheid: een email schrijven

SOLLICITATIEFORMULIER HOTEL PARKLUST

Voornaam: _____

Achternaam: _____

Straat + huisnummer: _____

Postcode + woonplaats: _____

Geboortedatum: _____

Telefoonnummer: _____

Op welke dagen wilt u werken? Zet 1 of meer kruisjes.

maandag
 dinsdag
 woensdag
 donderdag
 vrijdag
 zaterdag
 zondag

Op welke tijden kunt u werken? Zet 1 of meer kruisjes.

's morgens
 's middags
 's avonds

Welk werk wilt u doen? Zet 1 of meer kruisjes.

in de keuken
 in de bediening
 schoonmaken

Waarom wilt u dit werk doen?

Wat is uw werkervaring?

Datum: _____

Handtekening: _____

¹⁵ De voorbeelden van examenoefeningen van dit inburgeringsexamen zijn te verkrijgen via de site <https://www.inburgeren.nl/examen-doen/oefenen.jsp>. Van elk besproken examenonderdeel (behalve de oriëntatie op de Nederlandse arbeidsmarkt) is daar een oefenexamen beschikbaar.

Figuur 4.2: oefenopgave examenonderdeel schrijfvaardigheid: een formulier invullen

Dit examen wordt niet automatisch maar handmatig nagekeken. DUO biedt naast het oefenexamen een nakijkmodel voor elk van de drie onderdelen. Wat hierbij opvalt is de schijnbaar hoge mate van subjectiviteit die een beoordelaar geacht wordt te gebruiken bij het nakijken van de opdrachten. Zo worden de aanvulzinnen en de correspondentieopdracht beide beoordeeld op adequaatheid en begrijpelijkheid van de ingevulde tekst. Hiervoor kan een brief maximaal 3 punten krijgen, waarbij 0 punten gegeven worden wanneer de tekst “niet leesbaar en/of herkenbaar als uitwerking van de opdracht” is, 1 punt wanneer de tekst “nauwelijks acceptabel” is, 2 punten voor een tekst die “acceptabel en herkenbaar als uitwerking van de opdracht” is, en 3 punten voor een “volledig acceptabel” product.¹⁶

De beoordeling van grammatica gebeurt op dezelfde manier. Hier kunnen maximaal 2 punten behaald worden, waarbij van 0 tot 2 punten respectievelijk “veel”, “redelijk wat” en “nauwelijks of geen” grammaticale fouten gemaakt worden.¹⁷ Wat deze kwantitatieve oordelen precies inhouden, wordt in het beoordelingsmodel niet duidelijk gemaakt.

Voor spelling wordt 1 punt toegekend wanneer voldaan is aan het fonetisch ‘redelijk’ correct spellen van veelvoorkomende korte woorden en persoonlijke gegevens (wanneer hier niet aan voldaan wordt krijgt de deelnemer 0 punten), en 2 punten (het maximum) wanneer de spelling ‘redelijk correct’ is en de tekst begrijpelijk is.

Het beoordelen van woordenschat lijkt iets minder subjectief. Hiervoor staat 1 punt voor het gebruik van ‘standaardpatronen, uitdrukkingen en kleine groepen van woorden’ waarmee iemand beperkte informatie kan overbrengen, en 2 punten wanneer de woordenschat ‘redelijk gevarieerd’ is. Wat hier specifiek mee bedoeld wordt, wordt echter wederom niet duidelijk gemaakt.

Verdere criteria zijn coherentie in de briefschrijfopdracht (1 punt voor een duidelijke samenhang, 0 punten wanneer die er niet is) en het zich houden aan conventies bij het invullen van een formulier (1 punt voor correct ondertekenen, 0 punten voor niet of incorrect ondertekenen).

4.4.4 Examenonderdeel spreekvaardigheid

Het examenonderdeel spreekvaardigheid wordt afgenomen met behulp van een computer. Voor het voltooien van dit onderdeel heeft elke deelnemer 35 minuten de tijd. Het examen bestaat uit 2 delen. In het eerste deel verschijnt krijgt elke deelnemer via een kort filmpje een vraag gesteld. Een

¹⁶ <https://duo.nl/zakelijk/images/beoordelingsmodel-correspondentie.pdf>

¹⁷ <https://duo.nl/zakelijk/images/beoordelingsmodel-aanvulzinnen-inburgeringsexamen-schrijven.pdf>

voorbeeld uit het oefenexamen van zo'n vraag is: "Reist u vaak met de bus? Vertel ook waarom." en "Wat vindt u een leuk programma op tv? Vertel ook welk programma u niet leuk vindt." Na het beluisteren van elke vraag dient de deelnemer met behulp van een headset het gewenste antwoord in te spreken. In totaal bevat het examen 12 van zulke opdrachten.

Het tweede deel van het examenonderdeel spreekvaardigheid bevat meerkeuzevragen. Ook hier gaat elke vraag vergezeld van een inleidend filmpje. Bij de eerste meerkeuzevraag van het oefenexamen spreekvaardigheid zijn een man en een vrouw te zien. De vrouw, Lynn, zegt tegen Eric, de man: "Ik heb morgen een afspraak met onze baas. Heb jij ook een afspraak met haar?"

Na het filmpje volgt de vraag "Wat kan Eric het beste zeggen?", met daarbij drie mogelijke antwoorden: A) Fijn, dank je wel voor de moeite!, B) Goed, ik kan morgen om tien uur., en C) Nee, ik had gisteren al een gesprek. Elk mogelijk antwoord wordt voorgelezen en het juiste antwoord moet worden aangeklikt. In dit onderdeel staat niet zozeer uitspraak centraal, maar wel de juiste woordvolgorde, betekenis en passende context van de antwoorden.

De beoordeling van dit onderdeel is hetzelfde als beschreven in de vorige paragraaf wat betreft de beoordeling van adequaatheid, woordenschat en woordgebruik, grammaticale correctheid en coherentie van de antwoorden. Er zijn 2 extra beoordelingscriteria die betrekking hebben op spreken. Ten eerste wordt gekeken naar vloeiendheid. Geen vloeiende spraak krijgt 0 punten, 'enigszins vloeiend' 1 punt, en vloeiende spraak met enkel pauzes en herformuleringen bij langere uitingen krijgt 2 punten.

Uiteraard wordt uitspraak ook beoordeeld, met 0 punten wanneer alleen een geoefende luisteraar de uiting met veel inspanning kan verstaan, 1 punt wanneer de uiting door een geoefende luisteraar zonder inspanning redelijk verstaan wordt, en 2 punten wanneer de uiting goed te verstaan is, zowel voor geoefende als ongeefende luisteraar. Er worden redelijkerwijs geen punten afgetrokken voor spreken met een accent of het zo nu en dan verkeerd uitspreken van een woord.

4.4.5 Examenonderdeel luistervaardigheid

Het examenonderdeel luistervaardigheid wordt afgenomen met behulp van een computer. Voor het voltooien van dit onderdeel heeft elke deelnemer 45 minuten de tijd. Het examen bestaat uit het bekijken van een aantal korte filmpjes en het beantwoorden van 2 of 3 bijbehorende vragen. In totaal bestaat de toets uit ongeveer 25 meerkeuzevragen.

Een voorbeeld uit het oefenexamen¹⁸ bevat een filmpje waarin een man in een apotheek keelsnoepjes wil kopen, maar wordt aangeraden tabletjes of hoestdrank te kopen, en veel thee met

¹⁸ <http://www.oefenexamensduo.nl/portal/appmodules/examen/startexamen.ctrl>

honing te drinken. Hij koopt uiteindelijk tabletjes en zegt dat hij hoestdrank vies vindt. Deelnemers moeten vervolgens aangeven wat de man heeft gekocht. Er zijn vier mogelijke antwoorden: A) tabletjes, B) thee, C) hoestdrank, en D) keelsnoepjes. Naast het verstaan van het Nederlands wordt elke deelnemer hier dus ook geacht een simpel gesprek goed te kunnen volgen en begrijpen; omdat elke deelnemer geïnstrueerd wordt eerst de vraag te lezen alvorens het filmpje te bekijken, weet hij of zij wel ongeveer welke informatie uit het filmpje van belang is om het correcte antwoord te kunnen geven. Voor dit examen is geen beoordelingsmodel beschikbaar.

5 Discussie

Na in hoofdstuk 3 relevante achtergrondliteratuur per deelgebied van taalvaardigheid te hebben besproken, en in hoofdstuk 4 voorbeelden te hebben gegeven van de opdrachten die men bij de inburgeringsexamens kan verwachten, wil ik proberen in dit hoofdstuk de onderzoeksvragen (hoofdstuk 2) te beantwoorden.

De centrale onderzoeksvraag is, nogmaals, in hoeverre de inburgeringstaalexamens in de praktijk gefundeerd zijn op theoretische en/of empirische bevindingen uit de wetenschap. Onderzoeksvraag twee is of er in de literatuur aanwijzingen te vinden zijn die zouden kunnen zorgen voor verbetering van deze examens.

Om deze vragen te kunnen beantwoorden zou ik idealiter de beschikking hebben over zowel de inburgeringsexamens zelf, als over een wetenschappelijke verantwoording van de ontwerpers van deze examens, waarin zij aangeven welke wetenschappelijke bevindingen zij hebben gebruikt bij het ontwerpen van deze toetsen.

Hier ontstaat echter een probleem. Verzoeken om deze informatie bij enkele instanties (Bureau ICE, Cito) die meewerken aan het ontwerpen van de inburgeringsexamens levert niets op. De inburgeringsexamens worden in opdracht van het Ministerie van Sociale Zaken en Werkgelegenheid gemaakt, en van daaruit is de opdracht gegeven niets over de wetenschappelijke achtergrondinformatie naar buiten te brengen. Bij rechtstreekse navraag bij het Ministerie zelf is men ook extreem terughoudend. Verder dan het opnemen van een verzoek om informatie en enige persoonsgegevens, om op termijn op mijn vraag terug te komen, wil men niet gaan.

Hierdoor mist essentiële informatie die nodig is om mijn onderzoeksvragen goed te kunnen beantwoorden. Ik zal hier dus een andere benadering moeten kiezen, om de onderzoeksvragen zo mogelijk toch te kunnen beantwoorden. De enige relevante documentatie die wel voorhanden is, zijn de oefenexamens op www.inburgeren.nl.¹⁹ Deze zal ik daarom hieronder kort afzetten tegenover de in hoofdstuk 3 besproken literatuur. Zoals ook in hoofdstuk 3, zal wederom de meeste aandacht uitgaan naar de onderdelen spreekvaardigheid en luistervaardigheid. Die zullen dan ook het meest uitgebreid behandeld worden. Hiervoor is per deelvaardigheid een verzameling criteria uit de literatuur samengesteld. Daartegen zullen de oefentoetsen spreek- en luistervaardigheid afgezet worden.

¹⁹ <https://www.inburgeren.nl/examen-doen/oefenen.jsp>

Leesvaardigheid

De oefenexamens leesvaardigheid bestaat elk uit een tiental teksten waarbij in totaal 25 vragen gesteld worden. De vraagstelling heeft duidelijk een communicatieve insteek. Elke opdracht is opgesteld in het Nederlands. De toets bestaat uit meerkeuzevragen.

De doeltaal is hier duidelijk de voertaal. De meningen over het effect van deze aanpak zijn verdeeld. Voor sommigen is het de ideale situatie (Kwakernaak, 2011). Anderen (Alderson, 2000) menen dat met een vraagstelling in de T2 bij deelnemers van een lager taalvaardigheidsniveau het taalbegrip wellicht niet optimaal getoetst wordt. Dit kan de validiteit van de toets verminderen. Het opstellen van inburgeringsexamens in alle verschillende moedertalen van mogelijk deelnemers zou echter een zeer tijdrovende, kostbare, arbeidsintensieve zaak zijn.

Het gebruik van meerkeuzevragen zorgt, ten opzichte van open vragen, voor betere resultaten bij deelnemers van lagere taalvaardigheidsniveaus (Rahimi, 2007). Dit komt de validiteit van de toets ten goede.

Schrijfvaardigheid

De oefenexamens schrijfvaardigheid bestaan uit open opdrachten. Het gaat hierbij bijvoorbeeld om het schrijven van een brief of e-mail, het invullen van een formulier, en het afmaken van zinnen.

Bij deze open opdrachten is het helaas niet duidelijk welke beoordelingscriteria er exact gehanteerd worden. Ook deze informatie mag van het Ministerie niet openbaar gemaakt worden.

Het wordt dus niet duidelijk hoe bij de beoordeling van deze examens de invloed van beoordelaarseffecten geminimaliseerd wordt.

Opvallend is dat er geen gebruik gemaakt wordt van meerkeuzetoetsen. Uit de literatuur blijkt wel dat de meningen hierover sterk verdeeld zijn, al zijn er aanwijzingen dat meerkeuzetoetsing voor schrijfvaardigheid inzicht kan bieden in bepaalde deelvaardigheden van schrijfvaardigheid (De Gloppe & Willemsen, 2014). Het implementeren van meerkeuzeopdrachten zou wellicht meer inzicht bieden in de schrijfvaardigheidsontwikkeling van inburgeraars, wat bruikbaar is wanneer zij bijvoorbeeld niet de eerste keer slagen voor het examen.

Spreekvaardigheid

Uitspraak

Uitspraak wordt in de Toets Gesproken Nederlands automatisch beoordeeld.

Nauwkeurigheid

In de Toets Gesproken Nederlands wordt automatisch beoordeeld op vocabulaire en zinsbouw. Een groter vocabulaire en complexere zinsbouw kunnen wijzen op een hogere mate van nauwkeurigheid (en een hoger algemeen taalvaardigheidsniveau).

Vloeiendheid

Vloeiendheid wordt in de Toets Gesproken Nederlands automatisch beoordeeld.

Zinsbouw

Zinsbouw wordt in de Toets Gesproken Nederlands automatisch beoordeeld.

Vocabulaire

Vocabulaire wordt in de Toets Gesproken Nederlands automatisch beoordeeld (zie Nauwkeurigheid).

Intonatie

Uit de literatuur blijkt niet dat intonatie in de Toets Gesproken Nederlands beoordeeld wordt.

Grammatica

Uit de literatuur blijkt niet dat grammaticale kennis (aan de hand van bijvoorbeeld het gebruiken van correcte werkwoordsvervoegingen) in de Toets Gesproken Nederlands beoordeeld wordt. In het tweede deel van de oefentoets spreekvaardigheid wordt wel de kennis van woordvolgorde en het gebruik van de juiste voornaamwoorden getoetst.

Snelheid van productie

Uit de literatuur blijkt niet dat de snelheid van productie wordt meegenomen als factor bij de beoordeling van taalvaardigheid. Er staat echter wel een tijdslimiet van 35 minuten voor het afronden van de oefentoets.

Spontane spraak

De Toets Gesproken Nederlands bestaat uit drie onderdelen, namelijk het herhalen van zinnen, het kort beantwoorden van vragen, en het geven van antoniemen bij voorbeeldwoorden. Alleen het kort

beantwoorden van vragen vereist het produceren van spontane spraak, al is het de vraag in hoeverre deze vorm een goede indicatie geeft van de daadwerkelijke spontane, communicatieve spraakcapaciteiten van de leerling.

Benadering communicatieve werkelijkheid

De oefentoets spreekvaardigheid maakt gebruik van korte filmpjes waarin een man of vrouw de leerling een vraag stelt. Hierbij kan de leerling 'oogcontact' maken met de persoon op het scherm. Deze vorm van simulering van een gesprekssituatie werd aanbevolen in de literatuur. In het tweede deel van de oefentoets krijgt de leerling filmpjes te zien van korte gesprekken, waarbij meerkeuzevragen beantwoord moeten worden. In dit deel ligt de nadruk meer op woordvolgorde en contextuele passendheid van de antwoorden.

Gebruik scoringsrubrieken

De beoordelingsrubrieken bij de oefentoets die op de website van DUO te vinden zijn²⁰, zijn gebaseerd op oordelen over de acceptabelheid van bijvoorbeeld uitspraak, vloeiendheid, en grammaticale correctheid. Uit de literatuur blijkt dat het beter is voor leerlingen om in een rubriek helder te maken welke vaardigheden precies verlangd worden, zodat de leerlingen een duidelijk beeld hebben van wat er van hen verwacht wordt.

Luistervaardigheid

Interactie

Er is geen sprake van interactie. De enige interactie is die tussen de acteurs in de videobeelden.

Aanpassing moedertaalspreker aan T2-luisteraar

Dit is het geval. Er wordt gebruik gemaakt van korte, simpele zinnen, waarbij de tekst op normaal tempo en uiterst zorgvuldig wordt uitgesproken.

Gebruik shadowing/herhaling

Het gebruik van shadowing is hier in zoverre relevant dat een leerling bijvoorbeeld de vraagstelling voor zichzelf kan herhalen (de vragen kunnen overigens ook opnieuw worden beluisterd). Omdat er geen sprake is van interactie is het voor de leerder ook niet mogelijk gebruik te maken van selectieve shadowing.

²⁰ <https://duo.nl/zakelijk/images/beoordelingsmodel-spreekvaardigheid.pdf>

Authentieke input volgens input hypothesis

Er wordt in de oefentoets gebruik gemaakt van relatief eenvoudig Nederlands. Het is niet te zeggen in hoeverre de input voldoet aan de hypothese van Krashen (1981), omdat het taalvaardigheidsniveau per leerling zal verschillen. Het deel van de input dat nog buiten het bereik van de individuele leerling ligt, zal dus ook per geval verschillen.

Benadering communicatieve werkelijkheid

Er is geen sprake van directe interactie tussen luisteraar en spreker. Er wordt geen communicatieve situatie, zoals een gesprek, gesimuleerd. Dit is een van de gebreken van toetsing van luistervaardigheid met behulp van een computer. Wel worden de situaties in de videobeelden zo authentiek mogelijk gepresenteerd.

Lengte en inhoud fragmenten afgestemd op individu

De gebruikte geluids- en beeldfragmenten zijn erg kort en bevatten eenvoudig taalgebruik. De fragmenten zijn echter niet afgestemd op het individuele niveau van elke leerling.

Visuele input ondersteunt gesproken tekst

De visuele input bij de vragen ondersteunt de gesproken tekst, biedt voldoende context, en leidt niet af.

Zelfbeoordeling

Er is geen sprake van zelfbeoordeling.

Aanbod vraagstelling zowel gesproken als geschreven

Alle vragen die de leerling te horen krijgt, zijn ook uitgeschreven. Enkel de tekst in de geluidsfragmenten en videobeelden is niet uitgeschreven.

6 Conclusie

In het voorgaande werd gekeken naar de toetsing van deelvaardigheden van taal. Vanwege de beperkte ruimte en de relevantie voor inburgeraars, is ervoor gekozen de focus te houden op de deelvaardigheden spreek- en luistervaardigheid. Aan de hand van criteria uit relevante literatuur over deze vaardigheden en de manier waarop deze vaardigheden getoetst kunnen worden, zijn de oefentoetsen voor spreek- en luistervaardigheid die online te maken zijn bekeken.

De toetsen blijken deels aan te sluiten bij de bevindingen uit de literatuur. Zo wordt spreekvaardigheid getoetst op basis van een beperkt aantal deelvaardigheden (omdat het onwerkbaar is alle relevante facetten van spreekvaardigheid te toetsen) die herhaaldelijk in de literatuur terugkwamen, zoals vloeiendheid, nauwkeurigheid, en zinsbouw. In het geval van de luistervaardigheid is de input afgestemd op het algemene niveau dat verwacht wordt van de leerlingen die deze toetsen maken. Ook de lengte en inhoud van de gebruikte input (zowel de geluidsfragmenten als de audiovisuele hulpmiddelen) zijn afgestemd op leerlingen met een lager vaardigheidsniveau in het Nederlands.

De meest in het oog springende beperking die de toetsen van beide deelvaardigheden echter vertonen, is het gebrek aan interactie en simulatie van de communicatieve realiteit. Uit de literatuur blijkt dat interactie bijdraagt aan de productie van vloeiende, natuurlijke spraak. Door het voeren van gesprekken en het houden van debatten wordt niet alleen de spreekvaardigheid getraind, maar ook de luistervaardigheid. Daarnaast voelen leerlingen zich meer op hun gemak, wat de vloeiendheid van spraak ten goede komt, en kunnen meerdere personen tegelijkertijd werken aan hun vaardigheden. Wellicht worden deze methoden wel gebruikt in inburgeringscursussen, maar bij de toetsing ontbreken ze vrijwel volledig.

Een groot verbeterpunt voor de inburgeringstoetsing zou dus kunnen bestaan uit het verminderen van individuele toetsing met behulp van computers, en de aandacht te verleggen naar interactie. Er zou meer focus moeten komen te liggen op het ontwikkelen van communicatieve vaardigheden, in plaats van de leerlingen te beoordelen op zuiver talige vaardigheden zoals woordvolgorde en uitspraak. Dit sluit beter aan op de werkelijkheid, en bereidt de leerlingen beter voor op het gebruiken van de Nederlandse taal in de maatschappij. Ook het afstemmen van onderwijs en toetsing op de individuele behoeften van leerlingen zou betere resultaten kunnen bewerkstelligen. De haalbaarheid van dit alles is, gezien de grote belangstelling voor inburgering en een beperkt aantal docenten, tijd, en geld, echter zeer twijfelachtig.

Helaas ontbreekt in dit onderzoek de meest essentiële informatie. Natuurlijk is inburgering een overheidszaak. Toch blijven de exacte redenen voor de geheimhouding van informatie over de wetenschappelijke fundering van de in Nederland gebruikte inburgeringsexamens onduidelijk.

Deze situatie leidt er helaas toe dat het niet mogelijk is de taalvaardigheidsexamens voor de inburgering te analyseren. Er kunnen alleen uitspraken gedaan worden over de oefenexamens, niet over de daadwerkelijk gebruikte inburgeringsexamens. Ook over de wetenschappelijke achtergrond van die examens kunnen geen sluitende uitspraken gedaan worden. Dit zorgt er helaas voor dat er slechts beperkt antwoord gegeven kan worden op de onderzoeksvragen.

Hopelijk wordt deze informatie op enig moment in de toekomst wel openbaar, zodat er meer onafhankelijk onderzoek gedaan kan worden. Dit is nu onmogelijk, omdat dit onderzoek nu enkel lijkt te worden gedaan door instanties die direct betrokken zijn bij de ontwikkeling van deze examens.

7 Bibliografie

- Alderson, J.C. (1984). Reading in a foreign language: a reading problem or a language problem. In Alderson, J.C. & Urquhart, A.H. (Eds.), *Reading in a foreign language*. London: Longman, 1-24
- Alderson, J.C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Anastasi, A. 1988: *Psychological testing*. 6th edition. New York: Macmillan.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language testing*, 17(1), 1-42.
- Bachman, L.F. & Eignor, D.R. (1997). Recent advances in quantitative test analysis. In Clapham, C. and Corson, D., editors, *Encyclopedia of language and education*. Volume 7: Language testing and assessment. Dordrecht: Kluwer Academic, 227–42.
- Banerjee, J. & Luoma, S. (1997). Qualitative approaches to test validation. In Clapham, C. and Corson, D., editors, *Encyclopedia of language and education*. Volume 7: Language testing and assessment. Dordrecht: Kluwer Academic, 275–87.
- Barnwell, D.P. (1996). *A history of foreign language testing in the United States: from its beginnings to the present*. Tempe, AZ: Bilingual Press.
- Baten, L. et al., Nederlandse Taalunie & Raad van Europa. (2008). *Gemeenschappelijk Europees referentiekader voor moderne vreemde talen: Leren, onderwijzen, beoordelen*. Nederlandse Taalunie. (2008).
- Boer, C. den, Dijkhuizen, M., & De Gooijer, J. (2011). *Luister! Het communicatieve aspect van luistervaardigheid*.
- Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International journal of listening*, 14(1), 14-31.
- Camp, H. (2012). The psychology of writing development—And its implications for assessment. *Assessing Writing*, 17(2), 92-105.
- Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System*, 29, 1-14.

- Gelderen, A. J. S. van (1992). *De evaluatie van spreekvaardigheid in communicatieve situaties: globale beoordeling en gedetailleerde analyse van spreekprestaties van 11-en 12-jarigen*. Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam.
- Glopper, K. D., & Willemsen, A. (2014). Kunnen gesloten toetsen bijdragen aan de toetsing van schrijfvaardigheid?. *Levende Talen Tijdschrift*, 15(1), 31-38.
- Gysen, S., Kuijper, H., & Van Avermaet, P. (2009). Language testing in the context of immigration and citizenship: The case of the Netherlands and Flanders (Belgium). *Language Assessment Quarterly*, 6(1), 98-105.
- Hamp-Lyons, L. (2014). Three current, interconnected concerns for writing assessment.
- Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal* 83(3), 367-384.
- Hunt, K. (1965). Grammatical structures written at three grade levels. Urbana: NCTE.
- Hunt, K. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the society for research in child development*, 35 (1), 1-67.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct?. *Applied linguistics*, 29(1), 24-49.
- Janssen-van Dielen, A. (1992). *Zelfbeoordeling en tweede-taalleren. Een empirisch onderzoek naar zelfbeoordeling bij volwassen lezers van het Nederlands*. Nijmegen: SSN.
- Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The modern language journal*, 86(4), 546-561.
- Jong, J. H. de, Lennig, M., Kerkhoff, A., & Poelmans, P. (2009). Development of a test of spoken Dutch for prospective immigrants. *Language Assessment Quarterly*, 6(1), 41-60.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2), 130-144.
- Kessens, I. J., Aarts, O. A. J., (2013). *Onderzoek naar de verschillen tussen menselijke en machinale scoring van de Toets Gesproken Nederlands* (No. TNO 2013 R11811). TNO.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford University Press.

- Kwakernaak, E. (2011). Drie vernieuwingsgolven in het vreemdetalenonderwijs. *Levende Talen Magazine*, 98(7), 18-22.
- Lado, R. (1961). *Language Testing*. New York: McGraw-Hill.
- Lazaraton, A. 1992: The structural organisation of a language interview: a conversational analytic perspective. *System* 20, 373–86.
- Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly*, 1-16.
- Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75(2), 196-204.
- Luoma, S. (2004). *Assessing speaking*. Ernst Klett Sprachen.
- Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL quarterly*, 36(2), 173-190.
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741-760.
- Milanovic, M. & Saville, N. (1996). Introduction. *Performance testing, cognition and assessment*. Studies in Language Testing, Volume 3. Cambridge: University of Cambridge Local Examinations Syndicate, 1–17.
- Moskal, B.M. (2000). *Scoring Rubrics: What, When and How?* Practical Assessment, Research & Evaluation, 7(3). In a peer-reviewed electronic journal, pp. 1-7.
- Murphey, T. (2001). Exploring conversational shadowing. *Language teaching research*, 5(2), 128-155.
- Nienhuis, L. J. A. (1977). Het toetsen van spreekvaardigheid (SVO reeks 8). *Den Haag: Staatsuitgeverij*.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York, NY: Peter Lang.
- Oller, J.W. (1979). *Language tests at school*. London: Longman.
- Onna, B. van., & Jansen, C. (2006). Nederland talenland? Over de beheersing van Engels, Duits, Frans en Nederlands in Nederlandse organisaties. In B. Hendriks, H. Hoeken en P. J. Schellens (Red.), *Studies in Taalbeheersing*, 2, (pp. 237-249). Assen: Van Gorcum.

- O'Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language testing*, 19(1), 33-56.
- Pienemann, M., Johnson, J. & Brindley, G. (1988) Constructing an acquisition-based procedure for language assessment. *Studies in Second Language Acquisition* 10, 217–43.
- Piller, I. (2001). Naturalization language testing and its basis in ideologies of national identity and citizenship. *International Journal of Bilingualism*, 5(3), 259-277.
- Pullens, T., Ouden, H. D., Herrlitz, W., & Bergh, H. V. D. (2013). Kan een meerkeuzetoets bijdragen aan het meten van schriftelijke taalvaardigheid?. *Levende Talen Tijdschrift*, 14(2), 31-41.
- Rahimi, M. (2007). L2 reading comprehension test in the Persian context: Language of presentation as a test method facet. *The Reading Matrix*, 7(1), 151–156.
- Rijlaarsdam, G. C. W. (1982). *Beoordelen van discussievaardigheid: constructie van analytische schema's voor de beoordeling van discussieleider en discussiedeelnemers*. Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam.
- Richardson, C. P., & Scinicariello, S. G. (1989). Television technology in the foreign language classroom. *Modern technology in foreign language education: Applications and projects*, 43-74.
- Rost, M. (2002). *Teaching and researching listening*. London, UK: Longman.
- Schmidt-Rinehard, B. C. (1994). The effects of topic familiarity on second language listening comprehension. *The modern language journal*, 78(2), 179-189.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1, 199-215.
- Shohamy, E. (1997). Testing methods, testing consequences: are they ethical? *Language Testing* 14, 340–49.
- Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.
- Spolsky, B. (1997). The ethics of gatekeeping tests: what have we learned in a hundred years? *Language Testing* 14, 242–47.
- Stokking, K. (2000). *Ontwikkeling en beoordeling van schrijfvaardigheden. Aanwijzingen en suggesties voor docenten Nederlands rond het werken met het 'schrijfdossier' in de Tweede Fase*. Utrecht: ICO/ISOR.

Stronks, M. C. (2011). Verscherpte inburgeringstoetsen.

Tannen, D. (1989). Talking Voices: Repetition. *Dialogue, and Imagery in Conversational Discourse*, -
Cambridge, New York.

Tauroza, S., & Luk, J. (1997). Accent and second language listening comprehension. *RELCJournal*, 28,
54-71.

Thomassen, T., & Haquebord, H. (2017). Doeltaal is voertaal: ook bij toetsen?

Leesvaardigheidstoetsen op A2 en B1 niveau onderzocht. *Levende Talen Tijdschrift*, 18(2), 14-20.

Varonis, E. M., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation of
meaning. *Applied linguistics*, 6(1), 71-90.

Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in
L2 reading comprehension. *Journal of Research in Reading*, 25(1), 81-95.