

# Evalueren van het niveau van studenten van HBO-opleidingen met de Educational Level Evaluator (ELE)

Trudy Rexwinkel\*, Peter Verhoeven\*, & Albert Pilot<sup>^</sup>

\* *Edulevel Eindhoven*, <sup>^</sup> *Universiteit Utrecht*

---

**Samenvatting:** Het meten en evalueren van het niveau van een hbo-opleiding op valide en betrouwbare wijze is een lastig probleem waarvoor nog geen adequate oplossing is gevonden. Wel zijn er verschillende initiatieven genomen en instrumenten ontwikkeld. Deze bleken echter 'bureaucratisch' en/of 'ambigu' wat de betrouwbaarheid van de meting in de weg staat. Er zijn instrumenten nodig waarmee valide en betrouwbaar wordt gemeten 'wat studenten leren, en op welk niveau'. De onderzoeksvraag is: welk instrumentarium kan worden ontworpen waarmee het niveau van afgestudeerden van hbo-opleidingen valide en betrouwbaar is te evalueren en (nationaal en internationaal) is te vergelijken?

Om aan deze vraag tegemoet te komen is met behulp van de Design Research Methodology (DRM) het conceptuele ontwerp van de Educational Level Evaluator (ELE) tot stand gekomen en zijn daarmee empirische studies uitgevoerd. Er zijn twee varianten van de ELE ontwikkeld: een kwantitatieve en een kwalitatieve benadering. Beide benaderingen laten zien dat zij voldoen aan de criteria van representativiteit, validiteit en betrouwbaarheid.

---

**Trefwoorden:** opleidingsniveau, validiteit, betrouwbaarheid, kwantitatieve benadering, kwalitatieve benadering.

---



Corresponderende auteur: Trudy Rexwinkel - [trudyrxl@gmail.com](mailto:trudyrxl@gmail.com)



Artikel ontvangen [19-11-2019]; geaccepteerd [09-01-2020]; online gepubliceerd [01-09-2020].

## 1. Inleiding

De verantwoordelijkheid voor het niveau van een hbo-opleiding ligt doorgaans bij het HO-management. Deze managers bezorgden onder meer de volgende vraagstellingen:

*"De laatste jaren hebben wij veel geïnvesteerd in de bacheloropleiding. Nu willen wij laten zien dat het niveau van deze opleiding goed is. Hoe kunnen wij dat aanpakken?"* (Manager A).

Opleidingsmanager B wil de impact bepalen van de grote veranderingen in het werkveld van de bacheloropleiding. Deze veranderingen zijn wetenschappelijk onderbouwd en hebben belangrijke consequenties voor onze opleidingen. Met andere woorden: *"Hoe is zicht te verkrijgen op de kwaliteit van de aansluiting van onze afgestudeerden op het werkveld?"*.

Een Landelijk Overleg van Opleidingen wil een valide en betrouwbare vergelijking hebben van de gerealiseerde niveaus van de opleidingen die aangesloten zijn bij dit overleg. Manager C: *"Kan een niveauproject hierin voorzien"?*

*"Onze bacheloropleidingen werken nu een aantal jaren op basis van een nieuw profiel, dat nodig was vanwege internationale ontwikkelingen."* Wij willen graag een internationale vergelijking van opleidingsniveaus (Manager D).

### **1.1 Context en onderzoeksvraag**

Het vraagstuk van het niveau van opleidingen in het hoger onderwijs is vooral relevant sinds de invoering van de bachelor- en masterstructuur. Vanuit de kwaliteitszorg werd gevraagd het feitelijke niveau met kwantitatieve gegevens valide en betrouwbaar weer te geven.

Voor de oplossing van dit probleem zijn verschillende initiatieven genomen en diverse instrumenten ontwikkeld. De *Quality Assurance Agency for Higher Education* (QAA) in het Verenigd Koninkrijk publiceerde bijvoorbeeld *subject benchmark statements* (criteria voor vakken). De bedoeling was dat opleidingen deze zouden hanteren bij het beoordelen van de opleiding. Binnen het hoger onderwijs in het VK werden deze criteria echter beschouwd als 'onaanvaardbaar bureaucratisch' (Bellingham 2008).

Europese beleidsmakers ontwikkelden de Dublin Descriptoren die de niveaus van bachelor- en masteropleidingen beschrijven (Bologna Working Group 2005). Het was de bedoeling dat deze kwalificaties zouden worden toegepast in de Europese landen om internationale vergelijking van het opleidingsniveau mogelijk te maken. Internationale beleidsmakers gingen akkoord met de algemeen geformuleerde beschrijvingen. De verwoording in algemene termen biedt echter ruimte aan te veel interpretaties, waardoor een valide en betrouwbare internationale vergelijking van het opleidingsniveau ons inziens niet mogelijk is (Rexwinkel et al. 2012a).

Cosgrove (2011) geeft aan dat het probleem van het opleidingsniveau 'complex' is en Ewell (2010) stelt dat het 'ongrijpbaar blijft'. Daarnaast zijn er de accreditatie-organisaties die problemen ondervinden bij het uitvoeren van hun kerntaak van het evalueren en beoordelen. Harris-Huemmert (2008) stelt op grond van ervaring met deze organisaties dat het bepalen van het niveau diepgaande kennis vereist van het vakgebied van de opleiding, internationale oriëntatie en ervaring in het beoordelen.

Ook de volgende problemen worden in de literatuur benoemd: expertise, empirische evidentie en transparantie.

'Expertise in het beoordelen is niet aanwezig in accreditatiecommissies' (Sandahl et al. 2006 p.9). 'Subjectieve beoordelingen van de programma-inhoud worden waargenomen', terwijl de European Association for Quality Assurance (ENQA) 'systematische oordelen eist van het leren van studenten en het gebruik van empirische evidentie voor hun pogingen om de kwaliteit te verbeteren'. In de afgelopen jaren is de NVAO de toetsdeskundigheid van visitatiecommissies gaan verhogen en controleren en kijkt nu naar de beoordeling van eindwerken van studenten (meestal scripties en maar zelden praktijkproducten van professionals).

Transparantie: 'Zonder dit type van concrete evidentie zijn alternatieve pogingen om de kwaliteit van het hoger onderwijs te verbeteren, waarschijnlijk verspilling en ineffectief' (Konrad et al. 2007).

Het Project '*Assessment of Higher Education Learning Outcomes (AHELO)*' van de OESO begon een haalbaarheidsstudie (Nusche 2008) en concludeerde dat, in de meeste landen, beoordelingsresultaten ontoegankelijk zijn. Als instituten voor Hoger Onderwijs de beoogde uitkomsten expliciet en op meetbare wijze zouden specificeren, dan zou vergelijkbare beoordeling van leerdoelen haalbaar zijn. Maar in 2015 constateerde Altbach (2015) dat dit AHELO-project een mislukking was geworden, dat het moeilijk was om te zien hoe een vervolgproject betere resultaten zou kunnen opleveren. Een vervolgproject is er op aanbeveling in 2012 van het *Institutional Management in Higher Education (IMHE board)* niet gekomen.

Er is dus volgens diverse onderzoekers en instanties nog steeds (Altbach, 2015) behoefte aan instrumenten die gericht zijn op het bepalen van het niveau, waarmee vergeleken kan worden wat studenten leren en die niet administratief van aard zijn. Het doel van dit onderzoek is zo'n instrumentarium te ontwerpen. De onderzoeksvraag is daarom: Welk instrumentarium kan worden ontworpen waarmee het niveau van afgestudeerden van HBO-opleidingen valide en betrouwbaar is te evalueren en te vergelijken?

### **1.2 Validiteit, betrouwbaarheid en representativiteit**

Uitspraken en conclusies over het opleidingsniveau dienen altijd gebaseerd te zijn op valide onderzoek. Daarvoor moet de bias worden herkend en geminimaliseerd. Validiteit betekent dat feitelijk wordt gemeten wat conceptueel beoogd wordt. Bias verwijst naar de aanwezigheid van storende factoren die de zuiverheid van de niveaumeting bedreigen. Storende factoren door bias bij een niveaumeting kunnen zijn: effecten van de

interviewer, de procedures, de wijze van organisatie, de mate van bekendheid met de methoden enz.

Bij niveau-onderzoeken zijn er vele mogelijke oorzaken voor bias. In dit artikel geven wij aan dat de bias tot een minimum kan worden teruggebracht met een zorgvuldig ontwerp van procedures, gebaseerd op een helder concept van het opleidingsniveau.

Betrouwbaarheid betekent dat een meting consistent het construct reflecteert dat wordt gemeten (Field 2005, p.666)

Een derde kritieke factor is de representativiteit. Dit heeft betrekking op de relatie tussen deelnemers aan het onderzoek (steekproef) en de populatie (in dit geval het werkelijke aantal studenten of afgestudeerden). Hierbij worden ook de kwalitatieve kenmerken van de deelnemers betrokken.

Voor de beantwoording van de hoofdvraag (het ontwerp en de evaluatie van het instrumentarium) is gebruik gemaakt van de *Design Research Methodology (DRM)*. Dit is een ontwerpgerichte onderzoeksmethodologie die activiteiten beschrijft met behulp van multidisciplinaire principes, terminologie en methoden (Blessing & Chakrabarti 2009). Met behulp daarvan is het conceptuele ontwerp tot stand gekomen, dat wij de *Educational Level Evaluator (ELE)* hebben genoemd. Vervolgens is dit instrument gebruikt bij pilotstudies. Op basis van die studies is tenslotte de waarde van het instrument geëvalueerd.

## **2. Het conceptuele ontwerp: het concept 'opleidingsniveau'**

Het concept 'opleidingsniveau' hangt samen met de missie en doelstellingen van de opleiding. In dit artikel ligt de nadruk op de mentale competenties die de student heeft ontwikkeld tijdens het leren in het hoger onderwijs en die de afgestudeerde beheerst bij zijn/haar functioneren in het werkveld, met name op het niveau van opleidingen in het hbo. Vermunt bestudeerde leerresultaten en concludeerde dat de leerresultaten konden worden gecategoriseerd als cognitieve, affectieve en regulatieve leerresultaten (Vermunt 1996; Martínez-Fernández & Vermunt 2013).

Eén van de belangrijkste doelen van bachelor- en masteropleidingen is dat studenten leren redeneren over (complexe) problemen in hun vakgebied, kortweg vakinhoudelijk redeneren. Studenten gebruiken cognitieve leeractiviteiten om inhoud te verwerken. Deze leggen de basis voor leerresultaten in termen van kennis, inzicht, en vaardigheden. Voorbeelden zijn: het zoeken naar relaties tussen onderdelen van het vakgebied en het gebruik daarvan in toepassingen.

Studenten verrichten affectieve leeractiviteiten om te leren omgaan met gevoelens die opkomen tijdens het studeren en die de voortgang van de leerprocessen positief of

negatief kunnen beïnvloeden. Voorbeelden: het zichzelf kunnen motiveren, het verbinden van subjectieve goedkeuringen aan leertaken, het beheersen van emoties die het leerproces belemmeren. De affectieve activiteiten van afgestudeerden hebben betrekking op emoties die de functie, professe of het beroep in essentie met zich meebrengen. Voorbeelden: voor de hbo-boekhouder is het essentieel dat hij/zij nauwkeurig werkt; 'nauwkeurigheid' is een essentieel affectief aspect van de professionele houding van de boekhouder. Dat is ook van toepassing bij verpleegkunde: op affectief vlak is het essentieel dat de verpleegkundige 'discreet' werkt aan bijvoorbeeld verpleegkundige-persoon relaties.

Studenten gebruiken regulatieve leeractiviteiten bij het organiseren en managen van hun cognitieve en affectieve leeractiviteiten, die indirect leiden naar positieve leerresultaten. Voorbeelden zijn het monitoren van de voortgang van een leerproces, het diagnosticeren van de oorzaak van moeilijkheden, en waar nodig het aanpassen van leerprocessen. De regulatieve activiteiten nemen een andere positie in bij het leren van studenten. Zij hebben een regulerende functie voor de cognitieve en affectieve leeractiviteiten en kunnen een bevorderende dan wel een belemmerende werking hebben op het niveau van studenten. De regulatieve activiteiten vormen dus een indirect onderdeel van het meten en vormen daarom geen onderdeel van het meetproces. Daarentegen komen zij voor als verklaring voor niveaoverschijnselen.

Voor een nadere precisering van de leerresultaten wordt gebruik gemaakt van twee taxonomieën. Voor het vakinhoudelijk redeneren wordt de SOLO-taxonomie van Biggs & Tang (2007 en 2015) gebruikt die staat voor de structuur van het geobserveerde leerresultaat (*Structure of the Observed Learning Outcome*) en die de leerresultaten van vakinhoudelijk redeneren in termen van complexiteit classificeert in vijf niveaus. Deze zijn aangegeven in tabel 1.

Tabel 1

*SOLO-taxonomie (aangepast van Biggs & Collis, 2014; Biggs & Tang, 2007, 2015)*

---

SOLO-niveau en beschrijving	
1	Op het <i>pre-structurele</i> niveau bevindt de student zich in de initiële fase van kennis van en inzicht in het nieuwe vakgebied. Hij/zij kan de kwintessens van een vraag of opdracht nog missen. Zijn/haar inzicht in de vraag is nog beperkt en om dit hiaat aan kennis toe te dekken, kan hij/zij tautologie toepassen.
2	Op het <i>uni-structurele</i> niveau, concentreert de student zich op een deel van de (domein) informatie waardoor de conclusie nog beperkt, mogelijk dogmatisch is. Hij/zij is in staat een aspect te herkennen, identificeren en definiëren.

- 3 Het *multi-structurele* niveau varieert van het oppakken (opnemen) van verscheidene onafhankelijke aspecten van de domeinkennis maar zonder deze uit te werken, tot het oppakken van vele aspecten die hij/zij toelicht. De student vertelt wat hij/zij weet (knowing-telling).
  - 4 Het *relationele niveau* is het eerste niveau dat relevant is voor het hoger onderwijs. De student is in staat delen van de domeinkennis met elkaar in verband te brengen en in een passende context te plaatsen. Hij/zij begint te denken als een professional die delen van het probleem integreert in een coherente structuur en die structuur gebruikt om een gegeven taak uit te voeren.
  - 5 Tenslotte, op het *uitgebreide abstracte niveau* is de student in staat declaratieve kennis in functionele kennis te transformeren (kennis die geleerd is tijdens de opleiding omzetten in kennis voor de beroepspraktijk). Hij/zij is in staat te theoretiseren, generaliseren en reflecteren over de grenzen heen van het relationele niveau en het vakgebied. Het coherente geheel wordt geconceptualiseerd op een hoger niveau van abstractie en wordt toegepast op nieuwe en uitgestrekte domeinen.
- 

De professionele houding vormt de affectieve component van het niveauconcept. Deze heeft betrekking op de essentiële, affectieve kenmerken van het beroep, domein of vakgebied waar de opleiding op voorbereidt. Voor het meten van de leerresultaten van de professionele houding construeerden wij de taxonomie voor de professionele houding (PH). Deze taxonomie voorziet in mogelijkheden om te indiceren hoe de houding van een lerende zich ontwikkelt in complexiteit bij het leren van de affectieve aspecten van een beroep, domein of vakgebied. Wij baseerden ons op de veel gebruikte taxonomie van Krathwohl, Bloom & Masia (1974) en op Zimmerman's (2006) cycli van zelfregulatie. Het resultaat is gegeven in tabel 2.

Tabel 2

*Taxonomie voor de professionele houding*

PH-niveau en beschrijving

---

- 1 *Attenderen*. Op het eerste niveau wordt de student zich bewust van bepaalde kenmerken van de professionele houding waarvoor hij/zij wordt opgeleid. Hij/zij wordt vanuit de opleiding geattendeerd op deze karakteristieken. Hij/zij voelt nog niet de noodzaak voor consistentie (samenhang) tussen de verschillende kenmerken.

## Evalueren van het niveau van HBO-opleidingen

- 2 *Accepteren.* De student accepteert de kenmerken van het beoogde gedrag en oriënteert zich daarop. Aanvullend wordt hij/zij geattendeerd op de cycli van *zelfregulatie*. Het leren van de aspecten van de professionele houding gebeurt via begeleide (geprogrammeerde) vormen van zelfregulatie. Hierbij komen aan bod: doelen, taken, geloof in eigen kunnen op professioneel gebied (*self efficacy*). Hij/zij focust op de beoogde verwachtingen aan de hand waarvan de student zijn/haar professioneel gedrag ontwikkelt en met behulp van zelfcontrole en zelfinstructie verbetert.
  - 3 *Goedkeuren.* Op het derde niveau keurt de student geleidelijk de kenmerken van de professionele houding goed (*valuing, waarden*). De afzonderlijke onderdelen worden waarneembaar in zijn/haar gedrag en geleidelijk aan meer consistent, hoewel de responses van de studenten nog inconsistent kunnen zijn met het beoogde gedrag.
  - 4 *Integreren.* De student laat zien dat hij/zij een aantal kenmerken van de professionele houding geïntegreerd en zonder inconsistenties kan toepassen
  - 5 *Internaliseren.* Op het hoogste niveau internaliseert de student verschillende kenmerken van een professionele houding wat erop wijst dat hij/zij de kenmerken van zijn/haar eigen gedrag consistent onder controle heeft.
- 

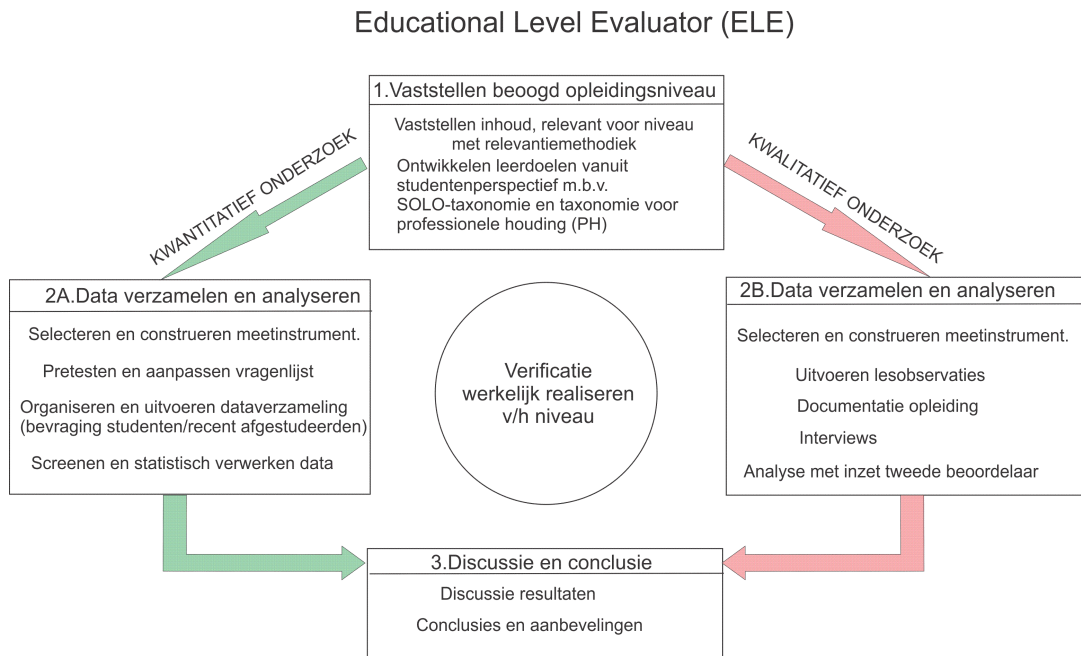
Zoals hiervoor is aangegeven nemen de regulatieve activiteiten een andere positie in bij het leren van studenten. Zij hebben een regulerende functie van de cognitieve en affectieve activiteiten en kunnen een bevorderende dan wel een belemmerende werking hebben op het niveau van studenten. Omdat de regulatieve activiteiten een andere positie innemen bij het leren van studenten, worden deze niet opgenomen in het meetsysteem. Wel kan worden aangenomen dat, door het bereiken van het eindniveau, de student heeft aangetoond te beschikken over voldoende regulatieve vaardigheden voor een adequate beroepsuitoefening.

### 3. Het Educational Level Evaluator (ELE)-instrumentarium

De ELE is een procedure met instrumenten die tot doel hebben het opleidingsniveau valide en betrouwbaar te evalueren en is gebaseerd op de Design Research Methodologie (Blessing & Chakrabarti, 2009). Het is een generiek model dat op diverse hbo-opleidingen kan worden toegepast. Figuur 1 (zie volgende bladzijde) geeft het overzicht van de stappen in het ELE-instrument weer.

Het is niet altijd mogelijk het opleidingsniveau vast te stellen met kwantitatieve, statistische methoden. Deze vereisen meestal grotere aantallen deelnemers waarbij bepaalde essentiële statistische analyses, zoals Principal Component Analysis (PCA)

toegepast kunnen worden voor het leveren van voldoende onderbouwing/bewijs van de validiteit. In de ELE wordt daarom naast de kwantitatieve methode ook de kwalitatieve methode beschreven. Het ontwerp van beide methoden wordt in dit artikel beschreven, evenals de evaluatie ervan in enkele hbo-opleidingen.



Figuur 1. Overzicht van de stappen in de Educational Level Evaluator

### 3.1 Vaststellen beoogd opleidingsniveau

Het doel van de eerste stap (gebaseerd op de eerder genoemde Design Research Methodologie) is het vaststellen van de inhoud die het meest relevant is voor het opleidingsniveau, in dit geval het gerealiseerde eindniveau van recent afgestudeerden. Dit gebeurt met de volgende sub-stappen:

- Creëren van thema's. Hiervoor hebben wij de relevantiemethodiek ontwikkeld, geïnspireerd door criteria uit accreditatieprocedures. Deze bestaat uit het creëren en bevestigen van inhoud/onderwerpen/thema's, resulterend uit de vraagstellingen:
  - Welke ontwikkelingen in het professionele veld en in de wetenschappelijke context zijn relevant voor het opleidingsniveau?
  - Welke ontwikkelingen in de externe omgeving zijn relevant voor het opleidingsniveau? De externe omgeving verwijst onder meer naar specifieke



wetgeving, beroepsorganisatie en sectorraad.

Tenslotte:

- Wat zijn de sterke onderdelen van het huidige curriculum? Elk thema wordt onderbouwd met wetenschappelijke literatuur.
- Bevestigen van thema's. Doel van deze sub-stap is het faciliteren van de overeenstemming over de validiteit van de thema's. De gecreëerde thema's worden gepresenteerd aan de stakeholders, studenten, docenten en leden van het management. Op een vragenlijst geven de stakeholders de mate van relevantie aan die zij toekennen aan de afzonderlijke thema's. De schaal reikt van 'niet' tot en met 'zeer relevant'.
- Ontwikkelen van leerdoelen. Nadat de thema's zijn bevestigd, worden hieruit leerdoelen ontwikkeld. Deze worden geformuleerd vanuit het perspectief van de studenten en gecategoriseerd als beoogde leerdoelen. De SOLO-taxonomie (Tabel 1) en de taxonomie voor de Professionele Houding (Tabel 2) worden geraadpleegd voor het formuleren van het vereiste niveau. Het 'relationele niveau' van SOLO-taxonomie is het eerste niveau dat relevant is voor het eindniveau van bacheloropleidingen. Denk aan hogere denkvaardigheden. Het eerste niveau dat relevant is voor het eindniveau van de professionele houding is het 'integreren van meer dan één kenmerk zonder inconsistenties'.
- Valideren van leerdoelen. De leerdoelen dekken de thema's inhoudelijk af en reflecteren het beoogde niveau (mate van complexiteit). Het valideren van deze leerdoelen vindt plaats met het vaststellen van de interbeoordelaarsbetrouwbaarheid bij meerdere beoordelaars door de berekening van Cohen's kappa coëfficiënt. Deze geeft de mate aan van overeenstemming over inhoud en niveau van de beoogde leerdoelen (Bryman 2012, p.280).

Tegenwoordig hebben opleidingen zodanige profielbeschrijvingen tot stand gebracht dat deze een belangrijke bron kunnen zijn bij deze sub-stappen.

### **3.2a Kwantitatief onderzoek**

We beschrijven nu eerst de werkwijze in het ELE-instrument (Figuur 1) en daarna (in paragraaf IV) de pilot-studies van het onderzoek. In het kwantitatief onderzoek wordt eerst het meetinstrument geselecteerd, waarmee data zoveel mogelijk '*unbiased*' verzameld en geschikt gemaakt worden voor betekenisvolle analyses. Het gerealiseerde eindniveau van de opleiding wordt het meest 'zuiver' gemeten bij recent afgestudeerden. Zij hebben de opleiding verlaten, niet langer dan 1.5 jaar geleden. Deze groep kan het best bereikt

worden met een vragenlijst, die speciaal voor het opleidingsniveau wordt geconstrueerd. Daarbij worden de leerdoelen uit de vorige stap omgezet in items van de vragenlijst, die de inhoud de onderwerpen afdekt. Met behulp van taxonomieën worden de items op het juiste niveau geformuleerd. Per item wordt gevraagd de mate van beheersing aan te geven op een schaal van vijf. Verder worden vragen naar kenmerken van afgestudeerden toegevoegd, zoals jaar van afstuderen, studievariant, positie van de afgestudeerde, en het type organisatie waar de afgestudeerde werkzaam is.

Ten behoeve van de kwaliteit van de vragenlijst wordt allereerst een concept-vragenlijst gepretest bij vierdejaars studenten.

De verkregen data worden geanalyseerd op representativiteit, validiteit en betrouwbaarheid. Eerst wordt inzicht verkregen in de representativiteit: het aantal deelnemers (steekproef) in relatie tot de werkelijke aantallen (populatie). Tevens worden de data gescreend op kwalitatieve kenmerken, bijvoorbeeld: zijn de recent afgestudeerden inderdaad niet langer dan 1.5 jaar geleden afgestudeerd. Tenslotte wordt de constructvaliditeit bepaald; deze statistische analyse wordt uitgevoerd met *Principal Component Analysis (PCA)*, die nagaat welke (groepen van) onderwerpen als de belangrijkste onderwerpen naar voren komen. De meting naar de betrouwbaarheid is gericht op interne consistentie van de belangrijke onderwerpen. De betrouwbaarheid van de meting is aangegeven met Cronbach's alpha coëfficiënt.

Wanneer de data voldoende representatief, valide en betrouwbaar blijken te zijn, worden deze verder statistisch geanalyseerd met SPSS, resulterend in gemiddelde scores en de spreiding daarin.

### **3.2b Kwalitatief onderzoek**

In het kwalitatief onderzoek wordt het niveau-concept van de opleiding(en) geoperationaliseerd zoals hiervoor beschreven onder 'vaststellen van het niveau' (onder figuur 1) en gericht op het meetinstrumentarium van de kwalitatieve benadering.

De meetinstrumenten bij de kwalitatieve benadering zijn in het algemeen: lesobservaties, documentatie van de opleiding en interviews met studenten en docenten. Er wordt vastgesteld welke observaties en interviews leiden naar de meest valide uitkomsten.

Met de regel van Cicchetti (1976) wordt het aantal analyses berekend, dat minimaal nodig is voor voldoende representativiteit van het onderzoek. Dat aantal moet minimaal  $2n2$  zijn, waar  $n$  het aantal kwalitatieve categorieën is (p.115). In het geval van het 'opleidingsniveau' heeft  $n$  betrekking op de cognitieve en affectieve component, dus  $n=2$ .

Het eindniveau van een opleiding kan het meest 'zuiver' gemeten worden bij studenten die zich bevinden in het laatste studiejaar. Voor de lesobservaties worden die studieonderdelen geselecteerd die het meest bepalend zijn voor het niveau van de opleiding.

Speciale aandacht moet geschonken worden aan de validiteit omdat dat een lastig begrip is bij de kwalitatieve benadering. Er is geen methode waarmee duidelijk is aan te geven wat de validiteit van kwalitatieve data is. Het is daarentegen wel mogelijk maatregelen toe te passen, die de validiteit bevorderen, zoals bijvoorbeeld: heldere definiëring van de leerdoelen op het gewenste niveau met gebruikmaking van de taxonomieën, beduchtheid op bias (doordat men zich anders gaat gedragen in de onderzoekssituatie) en door het toepassen van meerdere, verschillende databronnen.

De data worden verzameld d.m.v. lesobservaties en de interviews. Deze worden auditief opgenomen en later getranscribeerd en geanalyseerd. Op basis van deze analyse kan de beoogde beoordeling van de opleiding plaatsvinden. De beoordeling van de cognitieve en affectieve component(en) wordt uitgedrukt in een cijfer, bijvoorbeeld van een driepuntschaal.

Inzet van een tweede beoordelaar. Vanwege de mogelijkheid van enige subjectiviteit en onvolkomenheid worden de analyses van het materiaal (zonder de beoordelingen) voorgelegd aan een tweede beoordelaar, een procedure die verbonden is met Cohen's kappa. De tweede beoordelaar dient enerzijds onafhankelijk te zijn van de opleiding, anderzijds voldoende op de hoogte te zijn om zelfstandig te kunnen oordelen. De oordelen worden statistisch met elkaar vergeleken. De mate waarin de twee beoordelaars het meer dan toevallig eens zijn met elkaar wordt uitgedrukt in de inter-beoordelaars-overeenstemming: Cohen's kappa. Voor een redelijke overeenstemming moet deze liggen tussen 0.4 en 0.6.

Op deze wijze kan een resultaat van het kwalitatief onderzoek verkregen worden, met toch een indicatie van de betrouwbaarheid.

### ***3.3 Conclusie en discussie van het ontwerp van het ELE-instrumentarium***

De resultaten van het onderzoek bestaan na de voorgaande stappen uit een valide en betrouwbare bepaling van de mate van beheersing van de diverse onderdelen van het curriculum op het vereiste niveau. Deze resultaten kunnen daarna gepresenteerd worden aan de instellingen en met hen besproken worden.

## 4. Pilotstudies

De beschrijving van de pilotstudies is hier summier gehouden vanwege de beschikbare ruimte; voor een uitgebreidere behandeling wordt verwezen naar de desbetreffende, uitgebreidere, publicaties over dit onderzoek.

### 4.1 Verpleegkunde

De studie naar het niveau van vijf verpleegkunde-opleidingen is gepubliceerd in Rexwinkel et al., 2012b. Hiervoor is de kwantitatieve methode gebruikt.

Er is overeenstemming bereikt over de inhoud die het niveauconcept vormde bij verpleegkunde. Zij zijn verdeeld over de volgende thema's: Wetenschappelijk verplegen, Verpleegkundig zorgplan, Verpleegkundige-persoon relaties, Verpleegkundig diagnosticeren, die beschreven zijn met behulp van de SOLO- en PH-taxonomieën en wetenschappelijke onderbouwing.

Voor het meetinstrument werden de thema's (leerdoelen) omgezet in items voor de vragenlijst.

De data over het gerealiseerde niveau zijn verzameld met de vragenlijst bij recent afgestudeerden, dat wil zeggen niet langer afgestudeerd dan 1.5 jaar geleden. 700 afgestudeerden waren uitgenodigd deel te nemen aan de studie, het aantal respondenten was hoog (na een aanmaning): 603 (86%). Screening van de data leidde ertoe dat 133 respondenten werden verwijderd: velen van hen waren langer dan 1.5 jaar geleden afgestudeerd en diverse vragenlijsten vertoonden te veel ontbrekende waarden. Het netto aantal respondenten is 470, voldoende voor de kwantitatieve representativiteit.

Ook de kenmerken van de afgestudeerden, zoals het type organisatie waar afgestudeerde werkzaam is, en de studievorm, waren voldoende verspreid over de vijf opleidingen waardoor werd voldaan aan eisen van representativiteit.

De componenten zijn gemeten op de schaal van betrouwbaarheid met Cronbach's alpha coëfficiënt. Deze zijn  $\geq 0.81$ , waarmee zij tegemoetkomen aan de norm van  $\geq 0.70$  voor metingen van groepen (Commissie over Test Aangelegenheden [COTAN, 2011]). De *Principal Component Analysis (PCA)* identificeerde de vier belangrijkste thema's (Wetenschappelijk verplegen, Verpleegkundig zorgplan, Verpleegkundige-persoon relaties, Verpleegkundig diagnosticeren). Deze werden bevestigd als bestaande constructen. De data van de totale dataset voldoen aan de criteria van constructvaliditeit.

Om te kunnen vaststellen in hoeverre het niveau is bereikt, is de norm (door ons) bepaald op 3.00 (van de vijfpunt-schaal). Op basis van de valide en betrouwbare uitkomsten zijn z-scores berekend. De scores waren vaak tussen 2.00 en 3.00, wat laag is. De zelfbeoordelingen waren dus vrij laag. De standaarddeviatie was vaak  $\geq 1$  wat wijst op

## Evalueren van het niveau van HBO-opleidingen

een grotere spreiding dan gewenst en aanleiding kan zijn tot ambigue interpretaties. Daarom berekenden we z-scores van het gemiddelde en SD. De norm stelden we laag, op  $\geq 3.00$  omdat het de eerste keer was dat het niveau was geconceptualiseerd met taxonomieën in vakinhoudelijk redeneren en professionele houding. Het verschil tussen de norm en het gemiddelde werd gedeeld door SD [ $\text{norm-gemiddelde}/\text{SD} = z$ ] Deze 'z' werd met de cumulatieve standaard-normale verdeling (de tabellen van Abramowitz & Segun, 1964) omgezet in percentages van de respondenten die het niveau hebben bereikt. De componenten voldoen aan de norm als meer dan 50% van de respondenten hoger dan 3.0 scoorden. Per component is berekend in hoeverre de norm is bereikt. Dat gebeurt met het statistisch gemiddelde, standaarddeviatie en met het percentage afgestudeerden dat voldoet aan de eisen van het niveau (Tabel 3).

Tabel 3

*Statistisch gemiddelde (M), standaarddeviatie (SD), percentages respondenten  $\geq 3.00$  voor de vier thema's bij de vijf opleidingen afzonderlijk en de totale steekproef verpleegkunde*

	Case A	Case B	Case C	Case D	Case E	Totale steekproef
<u>Thema 1 Wetenschappelijk verplegen</u>						
M	3.22	2.77	2.69	2.87	2.60	2.85
SD	1.03	0.96	1.05	0.94	0.98	1.01
%	58	41	38	44	34	44
<u>Thema 2 Verpleegkundig zorgplan</u>						
M	3.50	3.26	3.14	3.12	2.95	3.20
SD	0.89	0.94	0.94	0.94	0.91	0.91
%	71	61	56	55	48	59
<u>Thema 3 Verpleegkundige-persoon relaties</u>						
M	3.44	3.26	3.23	3.12	2.86	3.16
SD	0.88	0.93	1.10	0.87	0.97	1.03
%	69	61	59	56	44	56
<u>Thema 4 Verpleegkundig diagnosticeren</u>						
M	3.08	2.93	2.56	2.99	2.51	2.78
SD	1.05	0.96	1.13	1.00	1.06	1.03
%	53	47	35	50	32	42

Tabel 3 laat zien dat de meeste problemen met het niveau zich voordoen in het eerste thema: Wetenschappelijk Verplegen. Vier uitkomsten (Case B t/m E) bevestigen dat dit thema nog niet voldoende aanwezig is in het curriculum van deze opleidingen. In lijn met wetenschappelijk verplegen, doen zich ook problemen voor bij het niveau van verpleegkundig diagnosticeren. Deze thema's impliceren dat de handelingen bij het verplegen meer gestandaardiseerd en gedefinieerd moeten worden met wetenschappelijke termen, met het doel de verpleegkundige interventies nauwkeuriger aan te geven.

#### **4.2 Hotelmanager, internationaal**

De studie naar het niveau van vier bacheloropleidingen hotelmanager in vier Europese landen is in detail beschreven in Rexwinkel et al. 2017. Ook hiervoor is de kwantitatieve benadering gebruikt.

Over de inhoud van het niveauconcept is met behulp van de 'relevantiemethodiek' overeenstemming bereikt. Aanvankelijk waren zes thema's ontwikkeld. Later in het proces vielen er twee af, namelijk Internationalisatie en Klantgerichtheid (wegens te weinig gewicht ( $\leq .40$ ) bij de *Principal Component Analysis (PCA)*). De hoofdthema's zijn Professioneel Management; Hospitality Business Research (HBR); Leiderschap; Strategisch Management.

Voor het meetinstrument werden de thema's (leerdoelen) eveneens omgezet in items voor de vragenlijst.

Verzamelen en analyseren van data. 733 afgestudeerden waren uitgenodigd deel te nemen aan de studie. Het aantal respondenten was 535 (73%). 142 respondenten werden geëlimineerd: 107 bleken langer dan 1.5 jaar geleden te zijn afgestudeerd en nog eens 35 vulden de vragenlijst niet compleet in (30% of meer ontbrekende waarden). Het netto aantal respondenten was 393. Dit aantal was voldoende voor een kwantitatieve pilotstudie, waarin een betekenisvolle analyse van constructvaliditeit en betrouwbaarheid kon worden toegepast.

Validiteit. De volledige steekproef van de vier opleidingen werd geanalyseerd met de Principal Component Analysis (PCA). De vier geconceptualiseerde thema's (Professioneel Management, Hospitality Business Research, Leiderschap en Strategisch Management) werden gemeten en bevestigd als bestaande constructen. De data van de volledige steekproef voldeden aan de criteria van constructvaliditeit wat wil zeggen dat er voldoende bewijs was dat de inhoud van deze meting correspondeerde met de inhoud van het construct dat beoogd werd te worden gemeten.

De betrouwbaarheid is met Cronbach's alpha coëfficiënt geanalyseerd bij de thema's. Deze waren hoger dan 0.70, waarmee zij voldoen aan de norm voor metingen van groepen (COTAN 2011).

Op dezelfde wijze als in de eerste pilotstudie is per thema berekend in hoeverre het niveau is bereikt bij de diverse opleidingen. De resultaten zijn op gelijke wijze gepresenteerd.

De uitkomsten van deze internationale vergelijking geven aan dat studenten uit drie landen/opleidingen (Oostenrijk, Nederland, en België) de leerdoelen van het thema Professioneel Management hebben bereikt, terwijl studenten uit twee landen/opleidingen (Oostenrijk en Nederland) de leerdoelen van het thema Leiderschap/Leidend Management hebben bereikt.

De uitkomsten van de opleidingen in Nederland, België en Noorwegen geven echter een lager niveau aan voor het thema 'Hospitality Business Research'. Alle opleidingen bleven onder de norm op het thema 'Strategisch Management'.

De conclusie is dat de meeste niveauproblemen bij deze opleidingen meer gerelateerd zijn aan 'vakinhoudelijk redeneren' (Hospitality Business Research en Strategisch Management) dan aan de 'professionele houding' (Professioneel Management en Leiderschap). De opleidingen in Noorwegen en België hadden lage scores in deze pilotstudie. Deze uitkomsten zijn geverifieerd bij de desbetreffende instellingen. Het bleek dat een van de opleidingen zich, ten tijde van de studie, bevond in een proces van herstructurering, terwijl de andere opleiding meer praktijk-georiënteerd was dan de voorlichting over de opleiding indiceerde.

### **4.3 Creatieve Therapie**

Voor de opleiding Creatieve Therapie (CT) is de kwalitatieve benadering gebruikt, omdat de opleiding, gezien de aantallen, zich niet leende voor een kwantitatieve benadering. Dit is de eerste maal dat de kwalitatieve benadering is toegepast, deze pilot moet daarom als exploratief worden gekenmerkt.

De opleiding CT verzocht om een bepaling van het eindniveau van alleen de vakken "Methodiek" en "Medium". De opleiding is ingedeeld naar de richtingen: Beeldende Therapie (BT), DramaTherapie (DT) en muziekTherapie (MT). Elke richting heeft zijn specifieke vakken "Medium" en "Methodiek". Voor het vaststellen van de inhoud die het meest relevant is voor het niveauconcept is de profielbeschrijving (LOO, 2016) een belangrijke bron. Met behulp van de relevantiemethodiek is naar niveaubepalende onderwerpen gezocht: belangrijk was daarbij de recente paradigmaverschuiving van

ziekte en zorg naar gezondheid en gedrag, positieve benadering en hogere denkvaardigheden. Empathisch vermogen en zelfregulering zijn eveneens relevant.

Bij Methodiek Leren de studenten het vakinhoudelijk redeneren. Het Vak-Therapeutisch Communiceren Leren zij hoofdzakelijk bij het vak Medium.

Het eindniveau wordt het meest 'zuiver' gemeten bij studenten die zich bevinden in het laatste studiejaar met de meetinstrumenten lesobservaties, interviews met studenten en docenten en documentaties van de opleiding.

Ter oriëntatie op de vakken Medium en Methodiek zijn docenten Beeldende Therapie en DramaTherapie geïnterviewd. De data voor het Vak-Therapeutisch Redeneren zijn verzameld in observaties van lessen Beeldende Methodiek, Muziek- en Drama-Methodiek. De data voor Vak-Therapeutisch Communiceren zijn verzameld in lessen Beeldend Medium, Muziek- en DramaMedium. In totaal is 17¼ uur audiomateriaal verzameld. Daarnaast is documentatie van de opleiding, in het bijzonder de profielbeschrijving, geraadpleegd.

Analyse van de lesobservaties. De getranscribeerde data zijn in de vorm van cases geanalyseerd. Omdat studenten in het laatste studiejaar niet meer beschikbaar waren, is gebruik gemaakt van studenten in het derde studiejaar. Vanzelfsprekend dienden de normen voor de leerdoelen daarvoor te worden aangepast. De data zijn geanalyseerd met de criteria voor Vak-Therapeutisch Redeneren en het Vak-Therapeutisch Communiceren.

Representativiteit. Deze studie is betrokken op derdejaars studenten, die zijn geobserveerd in de lessen Medium en Methodiek, afkomstig van alle aanwezige richtingen/media: Beeldende Therapie, Drama- en MuziekTherapie. Omdat deze lessen verplicht waren, was de steekproef bij benadering gelijk aan het werkelijke aantal studenten: 15 studenten Beeldende Therapie; 9 studenten MuziekTherapie en 6 studenten DramaTherapie.

De interviews zijn gehouden met drie docenten Beeldende Therapie, twee docenten DramaTherapie, twee docenten en een gastdocent MuziekTherapie.

Validiteits-bevorderende instrumenten zijn consequent gebruikt, zoals aangegeven in de beschrijving van het ELE-instrument.

Beeldende Therapie domineert in deze dataverzameling en -analyse ; hier zijn de meeste cases beschikbaar. Vak-Therapeutisch Redeneren heeft hoge beoordelingen, in het bijzonder bij Beeldende Therapie. Bij Drama- en MuziekTherapie zijn de beoordelingen lager. Het Vak-Therapeutisch Communiceren, de affectieve component van het niveau, is nog geen systematisch onderdeel van het curriculum bij Drama- en MuziekTherapie.



Een proefbeoordeling van de data van de beschikbare cases door een tweede beoordelaar resulteerde in een Kappa van .66.

## 5. Conclusie en discussie

De Educational Level Evaluator (ELE) is ontworpen als instrumentarium waarmee het niveau van studenten en/of afgestudeerden van hbo-opleidingen in het Europese hoger onderwijs valide en betrouwbaar is te evalueren. De ELE is een procedure die in stappen leidt naar het valide en betrouwbaar evalueren van het niveau. Daartoe wordt in de eerste stap het concept 'opleidingsniveau' zodanig uitgewerkt, dat het goed kan worden gemeten.

Met het conceptueel ontwerp van de ELE zijn empirische studies uitgevoerd. In dit artikel kwamen twee kwantitatieve studies en een kwalitatieve exploratie aan bod. Deze studies geven aan dat toepassing van de ELE leidt naar een valide en betrouwbare evaluatie van het niveau.

Voor de kwantitatieve methode is een vragenlijst gebruikt waarmee studenten en/of afgestudeerden worden bevestigd. De vragenlijst is zodanig ontworpen en toegepast, dat de bevestiging een valide en betrouwbaar resultaat geeft. Uit de twee pilots blijkt, dat dat inderdaad het geval is.

Voor de kwalitatieve methode is het instrument van de lesobservaties een essentieel onderdeel. De validiteit van de kwalitatieve bepaling wordt voor een belangrijk deel bereikt met validiteits-bevorderende maatregelen, zoals een heldere definiëring van het niveau, en objectiverende instrumenten zoals de beide taxonomieën. Voor de omvang van de benodigde steekproef wordt de eerder genoemde regel van Cicchetti gebruikt. Eventuele bias van de beoordelaar wordt voorkomen door gebruik te maken van een, onafhankelijke, tweede beoordelaar. Daardoor kunnen de beoordelingen worden gemiddeld en kan de betrouwbaarheid daarvan worden vastgesteld, uitgedrukt in Cohen's kappa.

De doelstelling van het onderzoek was niet alleen te komen tot een betrouwbare beoordeling van het opleidingsniveau, maar ook tot de vaststelling bij welke onderdelen van het curriculum hiaten wordt vastgesteld. Dit geeft de betreffende opleiding de mogelijkheid na te gaan hoe die hiaten kunnen worden weggewerkt en aldus het niveau te verbeteren.

Wat betreft de vraagstelling van de managers uit de inleiding van dit artikel kunnen beide benaderingen tegemoetkomen aan de wensen van de managers A en B. De vraagstellingen van manager C en D kunnen beter worden beantwoord door de

kwantitatieve benadering omdat zij betrekking hebben op de vergelijking van meerdere opleidingen.

## Referenties

- Abramowitz, M. & Segun, I.A. (Ed.) (1964). *Handbook of mathematical Functions, probability Functions*, (Table 26.1). New York: Dover publications INC.
- Altbach, P.G. (2015). AHELO: The myth of Measurement and Comparability. *International Higher Education* 82, 2-3.
- Bellingham, L. (2008). Assurance and the use of subject level reference points in the UK *Quality in higher education*. 14(3).
- Biggs, J. & Tang, C. (2007). *Teaching for quality learning at university* [3rd edition] Maidenhead, Berkshire: Society for Research into Higher Education (Open University Press).
- Biggs J. & Tang, C. (2015). *Teaching for quality learning at university: What the student does*. [1st edition] Open University Press McGraw-Hill Education McGraw Hill house Birkshire, UK.
- Biggs, J.B. & Collis, F. (2014). *Evaluating the quality of learning. The SOLO Taxonomy*. New York: Academic Press.
- Blessing L.T.M. & Chakrabarti A. (2009). *DRM, a Design Research Methodology*. Dordrecht: Springer.
- Bologna Working Group (2005). *A framework for Qualifications of The European Higher Education Area*. Copenhagen: Ministry of Science, Technology and Innovation.
- Bryman, A. (2012). *Social research methods*. 4th edition. Oxford University Press.
- Cicchetti, D. V. (1976). Assessing interrater-reliabilities for rating scales: Resolving some basic issues. *British Journal of Psychiatry*, 129,452-456.
- Cosgrove, R. (2011). Critical thinking in the Oxford-tutorial. *Higher Education Research & Development* 30(3): 343-56.
- COTAN, Committee on Test affairs Netherland (2011). *System of assessing the quality of tests of Psychologists /NIP*. Betrokken van <http://www.psynip.nl>.
- Ewell, P. (2010). Twenty years of quality assurance in higher education. *Quality in Higher Education* 16(2):173-176.
- Field, A. (2005). *Discovering Statistics Using IBM SPSS-Statistics*. Los Angeles: Sage.
- Harris-Huemert, S. (2008). Evaluators in of higher education in Germany. *Quality in Higher Education*, 14(1): 55-66.
- Konrad, H.et al. (2007). *Report of the Committee for the review of the Accreditation Organisation of the Netherlands and Flanders (NVAO)*. European Association of Quality Assurance (ENQA). Betrokken van [http://www.enqa.eu/pubs\\_review.lasso](http://www.enqa.eu/pubs_review.lasso)
- Krathwohl, D., Bloom, B., & Masia, B. (1974). *Taxonomy of educational objectives. Handbook II: Affective domain*. New York, NY: David Mc-Kay.
- Landelijk-Overleg-opleidingen vak-therapeutische beroepen (LOO VTB) (2016). *Domeinprofiel bacheloropleidingen vak-therapeutische beroepen*. Nijmegen: LOO VTB.

## Evalueren van het niveau van HBO-opleidingen

- Martínez-Fernández, J.R., & Vermunt, J.D. (2013). A cross-cultural analysis of the patterns of learning and academic performance of Spanish and Latin-American undergraduates. *Studies in Higher Education*, 40(2), 278-295.
- Nusche, D. (2008). *Assessment of learning outcomes in higher education*. Organisation for Economic Co-operation and development (OECD), Directorate for Education. <http://www.oecd.org/dataoecd/13/25/40256023.pdf>.
- Rexwinkel, T., Haenen, J., & Pilot, A. (2012a). Evaluating the level of degree programmes in higher education: Conceptual design. *Assessment & Evaluation in Higher Education*. Doi:10.1080/02602938.2012.755150.
- Rexwinkel, T., Haenen, J. & Pilot, A. (2012 b). Evaluating the level of degree programmes in higher education: the case of nursing. *Assessment & Evaluation in Higher Education*. Doi:10.1080/02602938.2012.751962.
- Rexwinkel, T., Haenen, J., & Pilot, A. (2017). The validity and reliability of the cross-national comparison of degree programme levels in European countries. What have students learnt? *European Journal of Psychology of Education*, 32(4), 703-723.
- Sandahl, R., Waerness, M., Brennan, J., & Sjunesson, K. (2006). *Evaluation of the Swedish national agency for higher education*. European Association of Quality Assurance (ENQA). [http://www.enqa.eu/pubs\\_review.lasso](http://www.enqa.eu/pubs_review.lasso).
- Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: a phenomenographic analysis. *Higher Education*, 31, 25–50.
- Zimmerman, B.J. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. In K. Ericsson, N. Charness, P. Feltovich & R. Hoffman, (ed.) *The Cambridge handbook of expertise and expert performance* 705-42. New York: Cambridge University Press.