

Adviezen in verband met het opnemen van ‘evidence based’ of ‘science based’ criteria in lastenboeken.

Inleiding

Er is een grote nood aan meer evidence based werken. Zowel vanuit moreel-ethisch standpunt (voorzichtige uitspraken doen over mensen omdat deze ernstige negatieve gevolgen kunnen hebben op het welzijn maar ook de socio-economische status van betrokkene en eventuele gezinsleden) als vanuit economisch standpunt (geen middelen verspillen).

Overheden hebben bovendien de verplichting om het zorgvuldigheidsbeginsel toe te passen. Ook al beschrijven wetteksten niet expliciet dat wetenschappelijke of ‘evidence based’ criteria dienen te worden gehanteerd, het lijkt ons van weinig zorgvuldigheid te getuigen indien men zich niet van wetenschap zou bedienen waar deze voorhanden is:

- Zo zou niemand het aanvaarden als de berekening van de draagkracht van een brug niet op basis van wetenschap werd uitgevoerd. Een brug die instort kan immers mensenlevens kosten.
- In de medische wereld wordt ook vooropgesteld dat artsen (maar ook klinisch) psychologen zich op wetenschappelijk onderzoek moeten verlaten. Stel u voor dat men u een wetenschappelijk bewezen effectieve behandeling zou ontzeggen ten voordele van onwerkzame kwakzalverij voor de behandeling van kanker.

Advies

Om het zorgvuldigheidsbeginsel gemakkelijker te respecteren en uit te voeren is het opnemen van *science en evidence based* criteria sterk aangewezen. Zij geven een veel grotere kans op objectiviteit dan subjectieve begrippen zoals ‘ervaring’ en referenties van tevreden klanten. Over die (on)betrouwbaarheid bestaat overigens op zich ook wetenschappelijk onderzoek. Een aantal overheden heeft er al voor gekozen om wetenschappelijke deugdelijkheid als criterium op te nemen. Zo vermeldt de omzendbrief BZ 2014/5 van de Vlaamse Overheid dat “*pseudowetenschappelijke praktijken (grafologie, astrologie, handlezing...) verboden zijn*” bij selectieprocedures (hier te consulteren:

https://overheid.vlaanderen.be/sites/default/files/Omzendbrief_BZ_2014_5.PDF)

Daarom stelden wij **voorbeeldteksten** op die u helpen om tegemoet te komen aan uw motivatieplicht. Dankzij deze teksten kunt u motiveren waarom u naar bepaalde informatie (échte bewijzen) vraagt en die u moeten helpen om een quoteringsysteem op te zetten dat een objectieve beoordeling garandeert.

Uitgangspunten

De uitgangspunten zijn eenvoudig gesteld de volgende:

1. Science Based
 - ① Deugdelijke **theorie**: is de theorie wel deugdelijk en aanvaard door de experts in het domein?
 - ② Evidence based: **empirie en statistiek**: werd er voldoende onderzoek uitgevoerd? hoe onafhankelijk is dat onderzoek? Hoe zat het methodologisch? Welke bewijzen leverde dit op?...
 - a) Wat zijn de empirische bevindingen?
 - b) Vragenlijst of test: werd deze deugdelijk ontwikkeld volgens geldende psychometrische normen?
2. **Publieke publicatie**: werden onderzoeken beschreven in het voor het publiek toegankelijke artikelen en minstens onderworpen aan **peer – review**? Zonder peer-review is er geen enkele controle of het beweerde juist is. Zelfs in de peer-reviewed academische wereld doen zich regelmatig schandalen voor (zie bijvoorbeeld retractionwatch.com), dus wanneer er geen peer-review is, is de kans op misleiding en frauduleuze bewering nog vele malen groter.

Wij adviseren u om de punten hieraan toegekend voldoende te laten meetellen in uw lastenboek om aldus eenvoudiger en betrouwbaarder tegemoet te komen aan het wettelijk verplichte zorgvuldigheidsbeginsel. Stel dat u een gewicht van 60% geeft aan het evidence based criterium in de totale beoordeling (naast de klassieke elementen zoals het criterium van de prijs dat bijvoorbeeld 30% krijgt), dan kan u vergelijken tussen aanbieders die wetenschappelijk betrouwbare producten (bijvoorbeeld selectietesten) of diensten (bijvoorbeeld individuele coaching) aanbieden.

Wij raden u ook aan voldoende expertise te organiseren om deze criteria te beoordelen – bijvoorbeeld door een beroep te doen op academici van een universiteit die expertise in het domein hebben.

De werkgroep ‘evidence based lastenboeken’.

Bart Van de Ven

Peter Catry

Patrick Vermeren

Michel Vandermeulen

(ps: Deze teksten werden ook nagekeken door de raadsman van de vzw Evidence Based HRM, meester Patrick Cockx)

Voorbeeldteksten die u kan opnemen in uw lastenboek/bestek/openbare aanbesteding

Wij verwachten dat u ons een aanbod maakt op basis van wetenschappelijk deugdelijke theorieën en praktijken op basis van de vier onderstaande criteria:

1. Theoretische deugdelijkheid

Maak een omschrijving van de wetenschappelijke status van de theorie waarop uw model of aanpak is gebaseerd en die u in dit dossier wenst in te zetten.

De theoretische deugdelijkheid telt mee voor maximaal XXX punten.

2. Bewijsvoering/empirische resultaten

Maak een omschrijving van de empirische onderzoeken die werden uitgevoerd. Vermeld het soort onderzoek, de methodologie en de resultaten. Vermeld de volledig referentie van elk document waarnaar u verwijst (naam of namen, publicatiedatum, titel, naam van boek of wetenschappelijk tijdschrift). Vermeld eveneens de impactfactor van de publicatie en het beoordelende orgaan (bijv. Thomson-Reuters).

De empirische bewijsvoering telt mee voor maximaal XXX punten.

3. In het geval van een vragenlijst of test:

Maak een omschrijving van de gebruikte methodologie en de resultaten bij de constructie van de vragenlijst of test. Vermeld de verschillende wetenschappelijke criteria die werden onderzocht en het resultaat. Vermeld de volledige referentie van elk document waarnaar u verwijst (naam of namen, publicatiedatum, titel, naam van boek of wetenschappelijk tijdschrift). De betrouwbaarheid en validiteit tellen mee voor maximaal XXX punten.

4. Peer-review (controle door academische vakgenoten):

Met betrekking tot theoretische deugdelijkheid en empirische bevindingen, kennen wij daarboven maximaal XXX punten toe, in functie van de impactfactor van het tijdschrift. Daarom vermeldt u eveneens de impactfactor van de publicatie en het beoordelende orgaan (bijv. Thomson-Reuters) in volgende tabel:

Naam Tijdschrift	Impactfactor volgens/ ranking in domein	Rangschikking	Gewicht
	Beoordelend orgaan: YYY		
		Top 10% binnen het vakgebied	10
		Volgende 15%	8
		Volgende 35%	6
		Volgende 20%	4
		Laatste 20%	2

(voorbeeld: u neemt het voorbeeld eventueel over)

Indien geen enkele kandidaat-leverancier minstens 50% behaalt van de maximum punten toegekend aan deze wetenschappelijke deugdelijkheid, zal de procedure worden stopgezet.

Voorbeeldteksten die u kan opnemen in uw motivatie van uw beslissing

Beoordeling van het criterium 'theoretische deugdelijkheid':

De theorie werd beoordeeld volgens een puntenschaal gaande van -6 tot +6. Uw theorie kreeg een beoordeling van ...

Beknopte uitleg: ...

Daardoor bedragen de punten voor dit onderdeel XXX op een maximum score van YYY.

Beoordeling van het criterium 'bewijsvoering/empirische resultaten':

De aangebrachte evidentie werd beoordeeld volgens een puntenschaal gaande van -4 tot +4. De door u aangereikte evidentie kreeg een beoordeling van ...

Beknopte uitleg: ...

Daardoor bedragen de punten voor dit onderdeel XXX op een maximum score van YYY.

(indien van toepassing) Beoordeling van het criterium 'constructie van de vragenlijst of test':

De door u aangebrachte artikels met betrekking tot de constructie van het meetinstrument werd beoordeeld op vlak van psychometrische betrouwbaarheid en diverse criteria inzake validiteit.

Beknopte uitleg: ...

Daardoor bedragen de punten voor dit onderdeel XXX op een maximum score van YYY.

Beoordeling van het criterium 'peer reviewed publicaties':

Op basis van de door u opgestelde tabel en onze controle hierom behaalde u een score van XXX op een maximum score van YYY.

Uw totaal aantal punten op theoretische deugdelijkheid en bewijsvoering bedraagt XXX punten op een maximumscore van YYY.

Wij concluderen op basis van het ingediende dossier dat...

Aanbevolen quoteringssysteem van de vzw Evidence Based hrm

Deze criteria werden opgesteld door Patrick Vermeren – copyright berust bij hem.

U kan de volgende graduele quotering toepassen:

① Deugdelijke **theorie**

- 6 meerderheid topexperten¹ noemen de theorie een bewezen 'theorem'
- 5 veel convergente bevindingen uit verschillende wetenschapsdomeinen
- 4 high level (meta)theorie: een brede verklaring voor een fenomeen
- 3 mid range theorie: een verklaring voor een bepaald fenomeen
- 2 low level (mini)theorie: verklaart een klein deelaspect van een fenomeen
- 1 theorie op basis van redeneren en logica/afgeleid van andere theorieën
- 0 er is geen theorie
- 1 methodologisch fout: 'dustbowl' of 'rainforest' empiricism/anecdotes
- 2 de theorie werd door academici verworpen of verlaten of staat haaks op valide/rationele theorieën
- 3 de theorie legt niets uit of is niet testbaar of men weigert ze te laten testen
- 4 theorie staat haaks op de bevindingen uit de fysica, biologie of chemie
- 5 vorige + maar men wil een impressie van wetenschappelijkheid wekken
- 6 men onttrekt zich met excuses aan de wetenschappelijke methode

Aan deze gradatie kunt u gewichten toekennen – dit is belangrijk van uit juridisch standpunt voor de motivatie van uw beslissingen. In de sociale wetenschappen zijn vele ondeugdelijke theorieën, dus dit moet het grootste gewicht krijgen, bijvoorbeeld 120 punten (bij een score 6 kent u 120 punten toe).

② Evidence based: **empirie en statistiek**

a) Empirische bevindingen?

- 4 geconfirmeerd in systematische review(s), bij voorkeur kwantitatieve reviews
- 3 onafhankelijk replicatieonderzoek bevestigt de resultaten
- 2 eigen replicatieonderzoek bevestigt de resultaten
- 1 de methode deugt maar er is nog maar één studie; er is nog onderzoek nodig
- 0 er zijn nog geen empirische bevestigingen
- 1 de gebruikte methoden zijn gebrekkig of niet adequaat
- 2 de evidentie is negatief of de vragenlijst is ondeugdelijk
- 3 meerdere studies verwerpen de theorie of de hypotheses
- 4 systematische reviews verwerpen de theorie of de theorie werd ondertussen ingetrokken

¹ Deze hebben publicaties in topjournals in hun vakgebied.

Aan deze gradatie kunt u gewichten toekennen – dit is belangrijk van uit juridisch standpunt voor de motivatie van uw beslissingen. U kan bijvoorbeeld 80 punten toekennen - een score van +4 komt overeen met 80 punten.

b) Vragenlijst of test: werd deze deugdelijk ontwikkeld? U kan volgende tabel met onderzoeksterminologie –en criteria gebruiken:

<p>Statistische betrouwbaarheid: meet de test of vragenlijst adequaat, zonder al te grote foutenmarge? Kan de meting herhaald worden met in grote mate hetzelfde resultaat? Opgelet: een goede betrouwbaarheid kan niet ingeroepen worden als bewijs voor de theorie.</p> <ul style="list-style-type: none"> • sub: interne consistentie (interne betrouwbaarheid) van de schalen: Alle correlaties tussen de items in een construct (concept, schaal) worden weergegeven. De meeste gebruikte methode is het berekenen van Cronbach's alpha. Een minimumstandaard op schaalniveau is 0.70; 0.80 is een zeer grote interne consistentie. • sub: test-heretest betrouwbaarheid: in welke mate geeft de test of vragenlijst hetzelfde resultaat wanneer de test een tweede keer wordt afgenomen 4 tot 8 weken later? Correlaties dienen minimum 0.70 te bedragen, vanaf 0.80 is heel betrouwbaar. • sub: split half betrouwbaarheid: in dit geval wordt de test in twee geplitst: de resultaten van beide afnames moeten hoge correlaties vertonen. • sub: parallele betrouwbaarheid: twee testen of vragenlijsten worden ontwikkeld die allebei hetzelfde concept zouden moeten meten. De resultaten van beide lijsten moeten hoge correlaties vertonen (> 0.40)
<p>Construct of concept validiteit: meet de test werkelijk wat de theorie beweert? Een validiteitscoëfficiënt kan nooit hoger uitvallen dan een betrouwbaarheidscoëfficiënt.</p> <ul style="list-style-type: none"> • sub: convergente validiteit: vertoont de test of vragenlijst correlaties met andere meetinstrumenten waarmee hij volgens de theorie zou moeten correleren (maar toch een ander construct meet – indien het om hetzelfde construct gaat moet u kijken naar concurrent validiteit (zie criterium validiteit). <ul style="list-style-type: none"> ○ sub: congruente validiteit: meten twee instrumenten exact hetzelfde construct? • sub: discriminante validiteit: zijn de constructen (schalen) voldoende gedifferentieerd van elkaar (ze moeten dus zo weinig mogelijk correlatie vertonen met de andere schalen ; de correlaties moeten nul of erg laag zijn) of verschillend van andere constructen die niet door deze vragenlijst worden gemeten (bijvoorbeeld een intelligentietest meet iets anders dan de een persoonlijkheidsvragenlijst over zes grote karaktertrekken)
<p>Inhoudsvaliditeit: dit is een niet-statistisch type validiteit: experts beoordelen op een systematische manier of de items (vragen, stellingen...) een voldoende representatief staal vormen voor het domein dat zij bevragen. De items zouden het construct volledig moeten afdekken.</p> <ul style="list-style-type: none"> • sub: representatie validiteit: werd het construct wel 'geoperationaliseerd'/'vertaald' naar observeerbare metingen? • sub: face validity: opgelet, dit wordt wetenschappelijk niet aanvaard als een validiteitscriterium: dit gaat over het aspect of mensen zich 'op het eerste zicht' herkennen in de items. Daarom wordt het ook vaak false/faith validity genoemd. het risico op het Forer-effect is aanwezig (mensen aanvaarden het resultaat omdat ze denken dat het uitgaat van een autoriteit of betrouwbaar instrument).
<p>Criterium validiteit: dit geeft een aanwijzing of de meting consistent is met wat reeds bekend is over het fenomeen.</p> <ul style="list-style-type: none"> • sub: concurrent validiteit: vertonen de resultaten van dit (nieuwe) meetinstrument voldoende grote correlatie met de resultaten uit reeds aanvaarde meetinstrumenten? Een correlatiecoëfficiënt van 0.50 is het minimum om te verwachten. • sub: predictieve validiteit: kan het meetinstrument iets voorspellen – met name dat wat men wou voorspellen? Bijvoorbeeld: in welke mate voorspelt deze intelligentietest toekomstige werkprestaties. <ul style="list-style-type: none"> ○ sub: incrementele validiteit: dit is niet echt een maatstaf van validiteit, maar dit cijfer drukt uit in welke mate de resultaten van een meetinstrument informatie of voorspelbaarheid kan toevoegen bovenop een andere test of vragenlijst (bijvoorbeeld: hoeveel voegt consciëntieusheid toe bovenop intelligentie in het voorspellen van toekomstige werkprestaties).
<p>Interne validiteit: werd het onderzoeksopzet (research design) uitgevoerd op een zorgvuldige en strikte manier? Werden ook alternatieve verklaringen in overweging genomen (of niet – in dat geval is er misschien confirmatieneiging). Werden andere variabelen bekeken?...</p>
<p>Externe validiteit: zijn de resultaten van deze studie generaliseerbaar (bijvoorbeeld naar andere populaties)? Waarom? Wat met mogelijke bias en fouten?</p>
<p>Ecologische validiteit: is het onderzoek wel representatief voor de 'echte' wereld – bijvoorbeeld: zullen de resultaten in een laboratoriumsetting herhaald worden in de echte wereld?</p>

Aan hoe meer van deze criteria correct wordt beantwoord, hoe meer punten u kan toekennen. U kunt bijvoorbeeld volgende maximale gewichtsverdeling toekennen:

- Betrouwbaarheid – interne consistentie: 10 punten
- Betrouwbaarheid – test-hertest betrouwbaarheid: 20 punten
 - Alternatief: Betrouwbaarheid – split half methode: maximum 10 punten op een totaal van 20 (omdat test-hertestbetrouwbaarheid te verkiezen valt)
- Betrouwbaarheid – parallele betrouwbaarheid: extra 5 punten indien dit werd onderzocht en sterke correlaties werden aangetoond.
- Validiteit – convergente validiteit: 5 punten
- Validiteit – discriminante validiteit: 10 punten (zich onderscheiden van een ander meetinstrument is belangrijk)
- Validiteit - Concurrente validiteit: 5 punten bij een nieuwe test (om geen nadeel te hebben indien het de eerste test in zijn soort is) OF 5 punten bij een vergelijking met een reeds erkend instrument
- Validiteit - Predictieve validiteit: 15 punten (voor heel veel wetenschappers is dit het belangrijkste validiteitscriterium omdat een test iets moet kunnen voorspellen over werkelijk gedrag).
- Validiteit: 3 punten per extra onderzocht criterium: externe validiteit, ecologische validiteit, interne validiteit, representatie validiteit

Dit levert een maximum puntentotaal op van 82 punten.

Bijkomende bemerkingen:

1. het verdient voorkeur dat bovenstaande validiteits- en betrouwbaarheidsstudies zijn verschenen in peer-reviewed tijdschriften (zie hierna).
2. Omdat dit niet altijd mogelijk is (omdat tijdschriften deze niet altijd aanvaarden) kan worden gekeken naar organismes die testen/vragenlijsten beoordelen, zoals de Nederlandse Commissie TestAangelegenheden (COTAN) van het Nederlands Instituut voor Psychologen, het Buros Institute of Review uit de Verenigde staten...
3. Vertalingen van oorspronkelijke vragenlijsten dienen een correcte procedure te doorlopen: backtranslation, validatiestudies (bijvoorbeeld controle van construct, factoranalyse, MDS...), normen voor de lokale publicatie.
4. Indien het bovenstaande ontbreekt, dient men vaak terug te vallen op rapportage (bijvoorbeeld een technische handleiding) van de uitgever/ontwerper zelf, maar van enige controle op fraude of fouten is dan geen sprake.
5. Wanneer een organisatie geen technische handleiding kan voorleggen, noch data over de bovenstaande criteria, kent u geen punten toe.

2. **Publicatie:** onderworpen aan **peer – review?**

Of artikels rond de theoretische deugdelijkheid, empirische bevindingen of psychometrische kwaliteiten zijn verschenen in peer-reviewed tijdschriften verwijzen wij in eerste instantie naar de lijst van tijdschriften met peer review zoals opgenomen

op de website www.ecoom.be; Ecoom is het Vlaams Expertisecentrum Onderzoek en OntwikkelingsMonitoring. Dit expertisecentrum garandeert een objectieve beoordeling.

De enige peer reviewed journals zijn A1, A2 en A3 publicaties.

A1-publicaties:

- Beantwoorden aan de hiervoor vermelde criteria
- Bevatten meestal ongepubliceerd materiaal
- Zijn 'peer-reviewed' of het proces waarbij collega-onderzoekers van hetzelfde domein de publicatie evalueren op zijn wetenschappelijke waarde.
- De internationaal erkende tijdschriften hebben een 'ISSN'-code
- Psychologie-artikels vind je via de **ISI Web of Science** databanken
 - Science Citation Index
 - Social Science Citation Index
 - Arts and Humanities Citation Index
 - Het kan gaan om: artikels, systematische reviews zoals meta-analyses, proceedings, letters (snelle mededelingen omwille van hun bijzonder belang), en notes.

Deze A1-publicaties *kunnen* ook een verschillend gewicht krijgen, naargelang de impactfactor. Het toekennen van een gewicht is een methode die wordt toegepast voor de financiering van de Belgische universiteiten. De universiteiten gebruiken de ISI impact factor van Thomson Reuters, maar dit is betalend. Bij het vermelden van de impactfactor dient de toekennende organisatie te worden vermeld (bijvoorbeeld de impactfactor van ThomsonReuters, Scimago Lab...). In het vakgebied psychologie zijn duizenden tijdschriften, maar niet allen hebben een even goede reputatie.

Naam Tijdschrift	Impactfactor volgens/ ranking in domein Beoordelend orgaan: YYY	Rangschikking	Gewicht
		Top 10% binnen het vakgebied	10
		Volgende 15%	8
		Volgende 35%	6
		Volgende 20%	4
		Laatste 20%	2

Enkele voorbeelden: Impactfactoren van tijdschriften binnen het domein Applied Psychology

Naam Tijdschrift	Impactfactor volgens/ ranking in domein Beoordelingsorgaan: Thomson-Reuters	Rangschikking	Gewicht
Journal Management	6,071 ISI impact factor 2014 binnen domein 'applied psychology': 1/76	Top 10% binnen het vakgebied (91-100%)	10
Journal of Applied Psychology	4,799 ISI impact factor 2014 binnen domein 'applied psychology': 2/76		
Personnel Psychology	4,49 ISI impact factor 2014 binnen domein 'applied psychology': 4/76		
		Volgende 15% (77-90%)	8
Personnel Psychology	0,921 ISI impact factor 2014 binnen domein 'applied psychology': 45/76	Volgende 35% (41-75%)	6
		Volgende 20% (21-40%)	4
Gedrag & Organisatie	0,375 ISI impact factor 2014 binnen domein 'applied psychology': 69/76	Laatste 20%	2

A2-publicatie:

- Soms klasseert men systematische reviews hieronder, maar dit is niet meest gebruikelijk
- Artikel in een internationaal wetenschappelijk tijdschrift met peer review, dat niet inbegrepen is in (A1)

A3-publicatie: Artikel in een nationaal tijdschrift met peer review, dat niet inbegrepen is in (A1) of (A2)

A4-publicatie: Artikel in een tijdschrift niet inbegrepen in (A1), (A2) of (A3) (noot: zonder peer review)

B = books of boeken

B1-publicatie: wetenschappelijk boek als auteur of co-auteur, geen (doctorale) dissertaties of conferentie-abstracts.

B2-publicatie: Hoofdstuk in een wetenschappelijk boek als auteur of co-auteur, geen conferentie proceedings (zie P)

B3-publicatie: Boekeditor of co-editor, inclusief conference proceedings (zie hierna).

P = Proceedings

P1-publicatie: Proceedings zijn een collectie academische artikels die gepubliceerd werden binnen de context van een academische conferentie en die werden opgenomen in één van de ISI Web of Science databanken (zie hiervoor); het kan gaan om artikels, review articles, proceedings paper, maar ze zijn niet onder de rubriek A1 opgenomen.

C = conference

C1-publicatie: Artikels in proceedings van wetenschappelijke congressen, die niet inbegrepen zijn in (A1) of (A2) of (A3) of (P1) (volledige artikels met uitsluiting van abstracts, ongepubliceerde lezingen, posters, ...)

C3-publicatie: Conference – meeting abstract, ongepubliceerde lezingen, posters,...

D = dissertation

D1-publicatie: Doctoraatsverhandeling (valt ook soms onder C2)

V = Varia

V-publicatie: Alle andere publicaties zoals kranten- of magazineartikels, bibliografieën, biografisch items, boekbesprekingen, correcties, redactionele artikels, tentoonstellingsbesprekingen, lemmata, rapporten, tekstedities, theaterbesprekingen, enz.