

Valide ijkingstoetsen in Vlaanderen

Jolan Hanssens, Greet Langie en Carolien Van Soom

Samenvatting In zijn proefschrift *Validation of positioning tests for higher STEM education – An argument-based approach* onderzocht Jolan Hanssens (2024a) de validiteit van twee Vlaamse start- en ijkingstoetsen. Deze toetsen beogen studenten te helpen bij hun studiekeuze door feedback te geven over hun startcompetenties. Vanuit de argumentatiebenadering op validiteit werd onderzocht in welke mate de toetsen doen wat ze beogen te meten en hoe ze gebruikt kunnen worden om studiekeuze en remediëring te ondersteunen.

Op basis hiervan werden drie aanpassingen aan de toetsamenstelling doorgevoerd. Ten eerste werden tekstbegripsvragen geschrapt omdat zij geen voorspellende waarde bijdroegen en een negatief effect hadden op de perceptie van deelnemers. Ten tweede werd het deel wiskundevragen opgesplitst in standaardwiskunde en eenvoudige basiswiskunde. Die laatste kunnen risicostudenten identificeren. Ten derde werden wiskundevragen in studierichtingsspecifieke contexten toegevoegd om de relevantie van de toets duidelijk te maken en zo de vrijwillige gevolgen die deelnemers eraan geven te versterken.

Deze reflectie sluit af met vijf adviezen voor het beleid en de implementatie van ijkings- en gerelateerde toetsen over (i) toetsamenstelling, (ii) cesuurbepaling, (iii) consequenties afwegen, (iv) communicatie over voorbereiding, en (v) validering en verplichte deelname.

Trefwoorden ijkingstoets, toetsvaliditeit, cesuur, studiesucces, studiekeuze

Artikelgeschiedenis

Ontvangen: 27 november 2025

Geaccepteerd: 19 juni 2026

Online: 17 september 2026

Contactpersoon

Jolan Hanssens, jolan.hanssens@kuleuven.be

Over de auteur(s)

Jolan Hanssens behaalde in 2024 zijn doctoraat aan de ku Leuven met onderzoek beschreven in deze paper. Hij is momenteel onderwijsassistent bij het Centrum voor Methodologie Van het Pedagogisch Onderzoek (CMPO) van de ku Leuven; Greet Langie is gewoon hoogleraar aan de Faculteit Industriële Wetenschappen van de ku Leuven, waar ze natuurkunde doceert. Zij is verbonden aan Leuven Engineering and Science Education Center (LESEC) en onderzoekt feedback, levenslang leren, de inclusie van meisjes in STEM, de transitie van secundair naar hoger onderwijs en van hoger onderwijs naar werkveld; Carolien Van Soom is hoogleraar aan de Faculteit Wetenschappen van de ku Leuven, waar ze biologie en wetenschapscommunicatie doceert. Zij is voorzitter van LESEC. Zij onderzoekt predictoren van studiesucces, feedback en de transitie van secundair naar hoger onderwijs.

Copyright

© Author(s); licensed under Creative Commons Attribution 4.0. This allows for unrestricted use, as long as the author(s) and source are credited.

Context en doelstelling

In Vlaanderen hebben beginnende studenten de keuze tussen 617 erkende initiële bacheloropleidingen (Vlaamse Overheid, 2025a). Daartussen kiezen is niet eenvoudig, terwijl het wel vaak gaat om een levensbepalende keuze (Demulder et al., 2023; Germeijs & Verschueren, 2006; 2007; Plank & Jordan, 2001; Tett et al., 2017). Studiekeziers hebben de moeilijke opdracht om de match of mismatch te beoordelen tussen hun interesses en startcompetenties enerzijds, en de inhoud van de opleidingen en het verwachte niveau van startcompetenties anderzijds (Hernandez-Martinez et al., 2011; Rigby, 2017). Ook voor hogere onderwijsinstellingen geldt dezelfde kenniskloof: zij kunnen vaak niet

goed inschatten of een student past bij een studierichting. De toegang tot het Vlaamse hoger onderwijs is dan ook open voor alle studenten met een diploma uit het secundair onderwijs (op een paar uitzonderingen na). Bovendien zijn er geen centrale examens aan het einde van het secundair onderwijs, al zullen vanaf 2027 wel centrale toetsen worden ingevoerd waarvan scholen kunnen beslissen om ze te laten meetellen naar het eindresultaat van hun leerlingen (Vlaamse Overheid, 2025c). Zelfs met deze aankomende toevoeging, impliceert dit systeem nog altijd vrijheid en autonomie voor studenten, maar legt het ook de hele verantwoordelijkheid van de studiekeuze in hun handen.

Hoewel de toegang tot het hoger onderwijs dus vanzelfsprekend is, is succes hierin dat uiteraard niet. De heterogeniteit in leerniveau bij het begin van hogere studies is een uitdaging zowel voor de studenten als de onderwijsinstellingen (OECD, 2015; Schelfout, 2022). In 2025 haalde slechts 27,6% van de Vlaamse studenten die 3 jaar eerder aan hun bacheloropleiding begonnen waren hun diploma. In 2023 was dat nog 31% en in 2009 zelfs nog 36% (Vlaamse Overheid, 2025b).

Om toekomstige studenten te helpen hun startcompetenties in te schatten en te toetsen aan wat er in het hoger onderwijs wordt verwacht, organiseren Vlaamse universiteiten in de zomervakantie ijkkingstoetsen (en starttoetsen) (ijkkingstoets.be). Ijkkingstoetsen zijn studierichting-specifieke, cognitieve testen met een adviserend doel. Zij verschillen dus van bijvoorbeeld ingangsexamens waarbij slagen een toelatingsvoorwaarde is, en waarnaar al meer valideringsonderzoek gebeurd is (e.g. Hallia et al., 2025; Husbands et al., 2015). Deelnemers met hoge scores op een ijkkingstoets worden bevestigd in hun voorlopige studiekeuze en studenten met lage scores worden aangemoedigd vroegtijdig te remediëren of hun studiekeuze kritisch te herevalueren. Een ijkkingstoets beoogt dus wel om *low-stakes* te zijn voor de deelnemers. Hoewel het resultaat nooit bindend is, zijn er wel studierichtingen die deelname aan een ijkkingstoets verplichten als voorwaarde voor inschrijving. In dat geval spreekt men specifiek van starttoetsen. In het vervolg van deze paper blijven we voor de eenvoud echter de benaming ijkkingstoetsen hanteren als overkoepelende term. Momenteel zijn de meeste studierichtingen met een ijkkingstoets natuurwetenschappelijk van aard.

Het doel van dit onderzoek was de validiteit van twee zulke ijkkingstoetsen in kaart te brengen en potentiële verbeterpunten te identificeren. De eerste toets is die voor de bacheloropleiding Industriële Wetenschappen (IW) – een opleiding tot industrieel ingenieur- en de tweede is die voor de inhoudelijk gerelateerde bacheloropleidingen Chemie, Biologie, Biochemie en biotechnologie, Geologie, en Geografie (CBBGG). Bij aanvang van dit onderzoek bestond de ijkkingstoets IW uit wiskundevragen en tekstbegripvragen; de CBBGG-toets voegde hier nog chemievragen aan toe.

Validiteitsargument als theoretisch kader

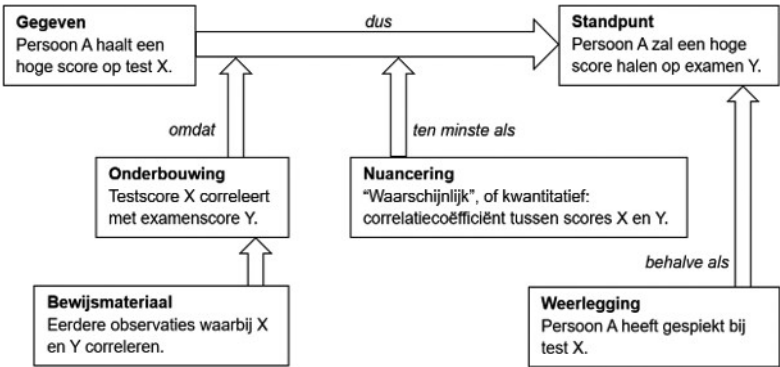
Het concept toetsvaliditeit is allesbehalve éénduidig gedefinieerd. *‘Meet een toets wat hij moet meten?’* heeft zelden een eenduidig antwoord (Kane, 2013). Zo kan men kiezen om validiteit te operationaliseren als inhoudsvaliditeit (Angoff, 1988; Sireci, 1998), criteriumvaliditeit (Cureton, 1951) (waaronder predictieve validiteit), of constructvaliditeit (Loevinger, 1957; Messick, 1989). Vaak wordt ook het geheel aan consequenties van een test mee geëvalueerd met validiteit (Folger et al., 2023).

Om de verschillende noties en bronnen van evidentie voor validiteit praktisch maar toch gestructureerd te hanteren, pleit Kane (2012, 2013) voor het gebruik van de logische argumentatiestructuur van Toulmin (1958) voor validiteitsonderzoek en zo een *interpretatie-gebruiksargument* voor de validiteit van een toets in een bepaalde context op te stellen, of kortweg een validiteitsargument. Een eerste kernpunt bij deze methode is dat validiteit geen eigenschap van een toets in se is maar wel van het gebruik van een toets.

Het tweede is dat het gebruik en de interpretatie van de toets beargumenteerd moeten worden op een gestructureerde manier. Veelal zal een testafname tot verschillende opeenvolgende inferenties leiden, welke dus allemaal beargumenteerd dienen te worden. Hoe die er exact uitzien, hangt van de context van de test af, maar enkele veelvoorkomende zijn: het scoren (de score is correct), het generaliseren van de score (de score weerspiegelt alle scores op gelijkaardige testen, i.e. de test is betrouwbaar), het extrapoleren (de score hangt samen met een interessant kenmerk, e.g. rekenvaardigheid), en de implicatie (de score wordt gebruikt voor iets, e.g. beslissen over slagen op een wiskundevak). De eerste drie inferenties zijn interpretaties en de laatste is een gebruik van de toets. Voor elk van deze hanteert Kane (2013) een Toulmin (1958) argument, bestaande uit 6 delen (figuur 1): een *gegeven* dat leidt tot een *standpunt*. Deze worden ondersteund door de *onderbouwing* en *bewijsmateriaal*. Een *nuancering* kan de geldigheid van het argument begrenzen en een *weerlegging* laat ruimte voor uitzonderingen.

Een derde belangrijk punt is dat niet elke interpretatie- en gebruiksinferentie empirisch aangetoond en beargumenteerd hoeft te worden. Aannames die heel waarschijnlijk zijn, kunnen a priori aanvaard worden, zodat gelimiteerde onderzoeksmiddelen aangewend worden voor de meer twijfelachtige aannames (Cronbach, 1989). Ten slotte is een validiteitsargument nooit volledig afgerond en is toetsvaliditeit dus altijd voorlopig (Kane, 2013).

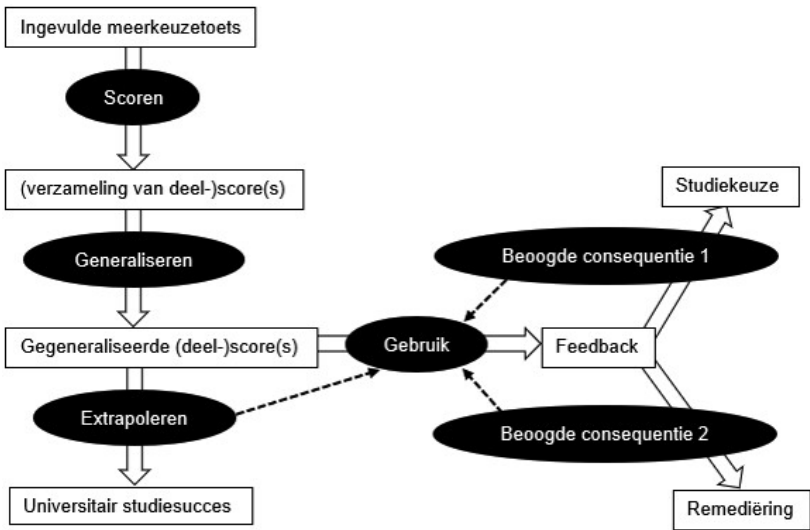
Valideringsonderzoek met een validiteitsargument is tweeledig. In een eerste stap structureert men het argument en in een tweede stap gaat men op zoek naar evidentie voor de onderbouwing. Het validiteitsargument voor ijkingstoetsen kan er als volgt uitzien. Een *scoringsinferentie* leidt van een ingevulde meerkeuzetoets naar een rechtvaardige testscore, of een batterij aan deelscores in het geval dat de toets uit meerdere delen bestaat. Deze scores zijn *generaliseerbaar* en dus betrouwbaar. De generaliseerde



Figuur 1 Schematische voorstelling van Toulmin (1958) argument met algemeen voorbeeld van een extrapolatie-inferentie

scores hangen samen met toekomstig studiesucces. De generaliseerbare scores worden *gebruikt* om studiekeizers feedback te geven. De studiekeizers geven hier zelf *gevolg* aan. Bij lage scores kunnen ze ofwel tekortkomingen remediëren, ofwel zelfkritisch hun eerdere studiekeuze herevalueren. Bij hogere scores voelen ze zich bevestigd in hun voorlopige studiekeuze. Figuur 2 schematiseert het validiteitsargument van ijkingstoetsen en wijst er ook op dat de *extrapolatie*- en beide *consequentie*-inferenties zullen dienen als onderbouwing voor de *gebruiksinferentie*. De implementatie van een ijkingstoets in de praktijk is immers enkel geoorloofd als de feedback die deze geeft echt gebaseerd is op samenhang met toekomstig studiesucces, en als studiekeizers de beoogde consequentie hechten aan deze feedback.

Kortom, in het kluwen van vele verschillende noties van toetsvaliditeit, biedt de argumentatiebenadering (Kane, 2012; 2013) een welgekomen praktisch houvast zonder de formele theorie uit het oog te verliezen. Het validiteitsargument als raamwerk helpt om mogelijke onderzoekspistes te structureren als verschillende inferenties vertrekkende van een afgenomen toets. Zo kan de validiteit van de toetsen holistisch en toch gefocust in kaart worden gebracht. Met name het idee om onderzoeksmiddelen te focussen op de meest twijfelachtige inferenties, biedt een richtinggevend houvast. Ook is dit raamwerk nuttig omdat ijkingstoetsen een Vlaams gegeven zijn en internationaal onderzoek ernaar bijgevolg schaars is. Het was bijgevolg niet mogelijk om de structuur en aanpak van een bestaand toetsvalideringsonderzoek toe te passen op de onderzochte situatie. Hoewel inferenties zoals “scores” en “generaliseren” veel voorkomen in validiteitsargumenten (Kane, 2013), vormen zij niet de focus van dit onderzoek, omdat zij niet als meest twijfelachtig worden beschouwd. Dit onderzoek focust op het vergaren van bewijsmateriaal voor en het verbeteren van de drie inferenties die de gebruiksinferentie rechtvaardigen en die a priori als twijfelachtig worden beschouwd, namelijk de extrapolatie en beide beoogde consequenties, zijnde het effect op studiekeuze en op remediëring.



Figuur 2 Validiteitsargument van ijkingstoetsen

Validiteitsargument en kernbevindingen

In de vorige sectie werd het validiteitsargument gestructureerd (figuur 2). Zoals besproken werd in dit onderzoek de focus gelegd op de extrapolatie- en beide consequentie-inferenties. In het vervolg van deze sectie worden kort de methodologie en vervolgens de voornaamste resultaten van het onderzoek besproken vanuit het perspectief van deze drie inferenties.

Methoden

Om deze interferenties te onderzoeken, werd gebruik gemaakt van een combinatie van kwantitatieve en kwalitatieve studies. Alle studies werden uitgevoerd bij zowel rw als CBBGG.

Om zinvolle voorspellingen van studiesucces te kunnen maken, moeten deze voorspellingen stabiel blijven over verschillende academiejaren. Alleen dan kunnen gegevens van eerdere cohorten gebruikt worden om uitspraken te doen over nieuwe cohorten, ook al veranderen zowel de toetsvragen als de samenstelling van de deelnemers van jaar tot jaar. Om deze stabiliteit te onderzoeken en te optimaliseren, werd de voorspellende waarde van twee manieren om ijkingstoetsscores uit te drukken met elkaar vergeleken: criterium-gerefereerde scores, geoperationaliseerd als de ruwe deelscores, en norm-gerefereerde scores, geoperationaliseerd als de percentielrang van een deelnemer binnen het cohort van afname (Hanssens et al., 2023b). Aan de hand van Chi²-testen werd vervolgens bepaald welke referentievorm retrospectief de meest stabiele classificaties kon opleveren.

De resultaten van de ijkingsstoetsen van meerdere cohorten werden vervolgens gekoppeld aan studievoortgangsgegevens uit het eerste bachelorjaar. Met lineaire en logistische regressie werd nagegaan in welke mate de verschillende toetsdelen studiesucces konden voorspellen, welke onderdelen daarbij de grootste voorspellende waarde hadden, en dewelke voorspellend waren bovenop elkaar en eerder gekende voorspellers zoals studiesucces in het middelbaar onderwijs.

Er werd ook een experimentele toetsafname in de eerste week van het academiejaar gedaan, waarbij deelnemers één toetsdeel kregen van deze die we met elkaar wilden vergelijken: naakte wiskunde, wiskunde in context, tekstbegrip, of chemie. Nadien volgde ook een vragenlijst (Smither et al., 1993) zodat ook de perceptie van deze toetsdelen kon worden vergeleken.

Het kwalitatieve luik van dit proefschrift bestond uit de thematische analyse van twintig semigestructureerde interviews bij ijkingsstoetsdeelnemers. Deze interviews brachten in kaart hoe studenten de toets en de feedback ervaarden, hoe zij de voorspellende waarde van de ijkingsstoets inschatten en in welke mate de feedback hun studiekeuze of remediëringgedrag beïnvloedde.

Extrapolatie

De extrapolatie-inferentie gebruikt ijkingsstoetsdeelscores als gegeven. Het standpunt is dat deze zinvol gebruikt kunnen worden om voorspellingen over studiesucces te maken. Om zinvolle voorspellingen te maken, is het van belang dat de voorspellingen stabiel zijn overheen verschillende academiejaren. Ondanks het huidige gebruik van criterium-refereren bleek norm-refereren stabielere voorspellingen op te leveren.

Onderbouwing en bewijsmateriaal voor deze inferentie liggen in de lineaire en logistische samenhang tussen ijkingsstoetsdeelscores en studiesucces (Hanssens et al., 2023b; 2024b; 2026). Hierbij werden vooral wiskundevragen en in iets mindere mate chemievragen (enkel voor CBBGG) als een goede voorspeller van later studiesucces geïdentificeerd. Tekstbegripsvragen bleken geen voorspellende waarde toe te voegen. Om studenten met een verhoogd risico op zwakke studieresultaten te identificeren werd de toevoeging van eenvoudige basiswiskundevragen onderzocht (Hanssens et al., 2024b). Deze bleken hier inderdaad voor geschikt. Studenten met lage scores hierop hadden dus een zeer groot risico op lage mate van studiesucces.

Hoewel het niet het expliciete doel van de basiswiskundevragen was, bleek ook dat de lineaire samenhang van de scores op deze vragen met studiesucces vergelijkbaar was met die van de standaardwiskundevragen. Ten slotte bleken ook contextuele wiskundevragen, ingebed in realistische en studierichtingsspecifieke contexten, voorspellend voor studiesucces, ook bovenop de andere toetsdelen (Hanssens et al., 2026).

De extrapolatie-inferentie wordt genuanceerd door kwaliteitskenmerken van de gevonden regressiemodellen en classificaties, zoals verklaarde variantie. Zoals vermeld waren deze voor de tekstbegripsvragen niet acceptabel, maar voor de andere deelscores wel. Kwalitatief onderzoek aan de hand van semigestructureerde interviews (Hans-

sens et al., 2023a) bracht ook een opmerkelijke weerlegging van deze inferentie aan het licht. Enkele studenten verwoordden potentiële criteriumcontaminatie. Hoog scorende ijkingstoetsdeelnemers zouden overmoedig worden en een lager studiesucces behalen, en laag scorende deelnemers zouden zich extra inzetten en een hoger studiesucces behalen. Deze stelling is tot op heden wel niet verder onderzocht.

Beoogde consequentie 1

De eerste consequentie-inferentie vertrekt vanuit het gegeven dat ijkingstoetsen gebruikt worden om deelnemers feedback te geven over de mate waarin ze de vereiste startcompetenties voor een bepaalde opleiding verworven hebben. Ze stelt dat deze feedback een effect kan hebben op de studiekeuze van deelnemers. De onderbouwing is dat deelnemers met positieve feedback een grotere kans hebben om in te schrijven in de corresponderende studierichting dan deelnemers met een negatieve feedback. Interviews leverden anekdotische evidentie op voor deze onderbouwing (Hanssens et al., 2023a). Bovendien bevestigen observaties en vragenlijstgegevens deze onderbouwing ook (Fonteyne et al., 2022; 2023). Men kan vermoeden dat een perceptie van relevantie van de toets voor de gekozen opleiding bij deelnemers deze inferentie versterkt (Hanssens et al., 2023a). In dit proefschrift werd aangetoond dat de eerder vermelde wiskundige contextvragen hiertoe bijdragen, en dat tekstbegripvragen de gepercipieerde relevantie kunnen ondermijnen (Hanssens et al., 2026). Als nuancering kan men stellen dat niet elke student de feedback ter harte neemt.

Beoogde consequentie 2

Ook de tweede consequentie-inferentie vertrekt vanuit hetzelfde gegeven dat ijkingstoetsen worden gebruikt om feedback te geven. Het standpunt is hierbij dat deze feedback kan leiden tot vrijwillige remediëring. De onderbouwing hiervoor is dat deelnemers met negatieve feedback een grotere kans hebben om te remediëren dan zij met positieve feedback. Ook hiervoor werd met de interviews anekdotisch bewijsmateriaal gevonden (Hanssens et al., 2023a), dat ook werd bevestigd door observaties en vragenlijstgegevens (Fonteyne et al., 2022; 2023). Zoals eerder vermeld, gaven de contextuele wiskundevragen de toets een perceptie van relevantie, namelijk dat de toets nauw aansluit bij de studierichting, en hadden de tekstbegripvragen het omgekeerde effect (Hanssens et al., 2026). Dit versterkt deze consequentie-inferentie, aangezien een relevantere perceptie kan leiden tot meer neiging om te remediëren wanneer de feedback dat vraagt (Fonteyne et al., 2023). Een nuancering is dat dit effect in principe enkel relevant is bij lage scorers met negatieve feedback. Uiteraard wordt er echter geïmpliceerd dat het voor een deelnemer die net geslaagd is, zinloos is om voorkennis op te frissen.

Implicaties voor onderwijsbeleid en -praktijk

Deze bevindingen ondersteunen het validiteitsargument van de ijkingstoetsen IW en CBBGG. Bovendien kunnen ze gebruikt worden voor het valideren van soortgelijke (ijkings- en start)toetsen. In wat volgt, worden implicaties van het gevoerde onderzoek voor het beleid en de toetspraktijk beschreven en worden adviezen geformuleerd.

Toetssamenstelling

Dankzij dit onderzoek werden drie verbeteringen aan de samenstelling van de ijkings-toetsen geïmplementeerd. Ten eerste werden de tekstbegripsvragen op beide toetsen geschrapt, omdat bleek dat zij noch de predictieve waarde noch de perceptie van deelnemers over de toets ten goede kwamen.

Ten tweede werden studierichting-specifieke contextvragen (wiskundevragen in een context die relevant is voor de corresponderende studierichting) toegevoegd, omdat die wel predictieve waarde hebben en de toets een hogere indruk van relevantie geven.

Ten derde werd het wiskundedeel opgesplitst in moeilijkere standaardwiskunde en eenvoudigere basiswiskunde. De standaardwiskundevragen hebben, naast hun predictieve waarde, als voordeel dat ze het verwachte leerniveau van wiskunde van een beginnende student weerspiegelen. De basiswiskundevragen zijn er vooral ter identificatie van risicostudenten. Deze weerspiegelen dus niet het verwachte startniveau. Een hoge score op deze vragen kan, met enige onzekerheid, gezien worden als een nodige maar geen voldoende voorwaarde op studiesucces in de latere opleiding. Een lage score kan anderzijds, weliswaar met gelijkaardige onzekerheid, worden geïnterpreteerd als een voldoende maar geen nodige voorwaarde op gebrek aan studiesucces.

Hieruit volgt het advies om de samenstelling van andere ijkingstoetsen (en andere gestandaardiseerde toetsen waarbij predictieve validiteit van belang is) kritisch onder de loep te nemen. Het is te verwachten dat eenvoudige wiskundevragen ook voor studiesucces in andere studierichtingen substantiële voorspellende waarde hebben voor risicostudenten. Wellicht is dit vooral het geval in opleidingen die enige wiskundige voorkennis veronderstellen, zoals bijvoorbeeld de bacheloropleiding Bioingenieurswetenschappen, maar ook menswetenschappelijke opleidingen zoals Economische Wetenschappen of Onderwijskunde.

Cesuren bepalen

Cesurbepaling is een kunst op zich. Vaak blijft het in de praktijk beperkt tot de helft van de maximumscore of gebruiken toetsontwikkelaars een informeel Angoff-protocol (1971) waarbij zij schatten hoeveel vragen correct beantwoord moeten worden om een bepaald advies te rechtvaardigen. Echter, om de validiteit van het gebruik en de interpretatie van de ijkingstoets te optimaliseren, wordt hieronder een optimalere methode aangeraden voor zowel de ijkingstoetsen IW en CBBGG, als andere ijkingstoetsen.

De voorgestelde methode steunt op het gebruik van normgerefeerde ijkingsstoetsdeelscores van voorbije academiejaren om stabiele voorspellingen te garanderen en op nutsfuncties om de waarde die aan een advies wordt gehecht te expliciteren. In een eerste stap wordt dan een ruwe individuele deelscore omgezet naar de percentielrang binnen het huidige cohort. Parallel wordt in een tweede stap een nutsfunctie gebruikt op beschikbare data van vorige cohorten. Men classificeert hier dan gemeten studiesucces op basis van percentielrang van ijkingsstoetsdeelscores in voorgaande cohorten. In zo'n nutsfunctie geeft men een waarde aan elk soort observatie (vals positief, echt positief, vals negatief en echt negatief) van een classificatie onder een bepaalde cesuur. Deze waarden worden opgeteld. De nutsfunctie kan vervolgens alle mogelijke cesuurwaarden doorlopen. Daar waar de nutsfunctie haar maximum bereikt, vindt men de optimale cesuur. Het gebruik van nutsfuncties is dus slechts beslissingsondersteunend. De uiteindelijke cesuurbeslissing moet steeds door een persoon genomen worden, maar deze persoon wordt gedwongen om bepaalde intuïties rond de waarde van verschillende observaties expliciet te maken. Een derde en laatste essentiële stap is dan simpelweg het vergelijken van de percentielrang van de score waarover men een feedbackbeslissing wil maken met de gevonden cesuur.

In een optionele vierde stap kan de communicatie naar de toetsdeelnemer gebeuren aan de hand van de vergelijking van de ruwe (criteriumgerefeerde) score vergeleken met de cesuur, ook omgezet naar een ruwe score. Dit sluit beter aan bij de terechte mening van veel studenten en ouders dat studiesucces niet afhangt van het leerniveau van het cohort, maar van dat van het individu. De cesuur zal hierdoor van jaar tot jaar verschillen, maar dat is voor deelnemers binnen hun cohort niet van belang.

Meerdere gebruiksinferenties en onbeoogde consequenties

Bij het begin van dit onderzoek hadden ijkingsstoetsen rw en CBBGG slechts één gebruik: niet-bindende feedback geven om studenten te helpen bij hun studiekeuze of de keuze om al dan niet vroegtijdig tekorten te remediëren. Sinds 2023 heeft de rw-toets echter een aanvullend gebruik: beslissen of studenten verplicht worden om te remediëren. Ook voor CBBGG wordt dit vanaf 2026 het geval.

Daarnaast tonen de interviews dat deelname aan de ijkingsstoets ook het beeld van studenten over de opleiding kan beïnvloeden. Hoewel dit niet het oorspronkelijke doel was (Vandewalle & Callens, 2013), kan het beschouwd worden als een bijkomend gebruik van de toets: studenten laten ervaren wat een opleiding inhoudt. De contextvragen kunnen hierin een cruciale rol spelen. Ze maken de link met het studieprogramma concreter en kunnen zo bijdragen aan een realistischer beeld van de opleiding. Zelfs als dit effect niet expliciet als doel wordt erkend, blijft het belangrijk dat de toets geen verkeerd of ontmoedigend beeld oproept bij studenten die wél geschikt zijn.

Ijkingstoetsen worden doorgaans beschouwd als *low-stakes*, omdat het resultaat formeel geen invloed heeft op toegang tot de opleiding. Uit de interviews blijkt echter dat dit beeld genuanceerder moet. Sommige studenten laten hun studiekeuze afhangen van hun

toetsresultaat, terwijl anderen dat niet doen. Dit is uiteraard de verwachting en ook een beoogde consequentie. Deze verschillen in ervaren inzet impliceren echter dat de toets niet voor iedereen dezelfde gevolgen heeft, ook niet bij gelijke scores, wat vragen oproept over de eerlijkheid van het ijkingstoetsproces. Uiteindelijk geldt dit als bijkomend argument om ijkingstoetsen optimaal te valideren, en met name het feedbackproces en hoe dat wordt ervaren te onderzoeken.

Het hierboven beschreven nieuwe gebruik, impliciete gebruik en verschil in ervaring illustreren het belang van het in kaart brengen van alle gebruiken en gevolgen van toetsdeelname als eerste stap van toetsvalidering in het algemeen. De negatieve gevolgen dienen te worden geminimaliseerd en de positieve gemaximaliseerd. Er dient een gegronde keuze te worden gemaakt over welke gevolgen er gekwantificeerd kunnen of moeten worden en vervolgens kunnen deze worden afgewogen voor een keuze omtrent implementatie van de toets (Sackett et al., 2010).

Communicatie omtrent voorbereiding

Hoewel op de website van de ijkingstoets (ijkingstoets.be) voorbeeldvragen beschikbaar zijn, wordt niet duidelijk gecommuniceerd of en hoe studenten zich moeten voorbereiden. Hierdoor ontstaat een heterogene groep deelnemers: sommigen bereiden zich voor, anderen niet. Vanuit validiteitsoogpunt is dit suboptimaal, omdat zo niet bij iedereen het competentieniveau in dezelfde omstandigheden wordt gemeten. Voor het oorspronkelijke gebruik van de ijkingstoets, niet-bindende feedback, hoeft dit geen probleem te zijn, zolang in de feedback vermeld wordt dat de interpretatie van resultaten afhangt van de mate van voorbereiding. Met de toevoeging van de verplichte remediëring voor lage scores is het echter wenselijker om het voorbereidingsniveau van deelnemers meer te uniformiseren. Uiteraard zal, ongeacht de boodschap (“wel” of “niet” voorbereiden), altijd een deel van de studenten zich toch niet op dezelfde manier hebben voorbereid. Aangezien elke vorm van cognitief toetsen onvermijdelijk ook een zekere mate van voorbereiding meet, is het veiligste en meest consistente uitgangspunt om studenten expliciet aan te moedigen zich voor te bereiden. Dit sluit bovendien aan bij het doel om laag scorende studenten effectief te begeleiden naar remediëring. Dit advies geldt voor alle ijkingstoetsen.

Verplichting en validering

Bij de aanvang van dit onderzoek was deelname aan ijkingstoetsen niet verplicht. Sindsdien is dat veranderd en gebruiken veel studierichtingen de ijkingstoets als starttoets. Bij een lage score mag men nog steeds inschrijven, maar deelgenomen hebben aan een ijkingtoets is dan wel verplicht. De uitzonderingen hierop zijn CBBGG (vanaf 2026–2027 is verplichting wel het geval) en de bachelors in de Artificiële intelligentie, Computerwetenschappen en Informatica. Het is een begrijpelijk standpunt van opleidingen dat ze eerst garanties rond de validiteit van de toets willen en deze dan pas willen verplichten.

Een verplichte afname van de toets zou echter noodzakelijkerwijs leiden tot grotere datasets voor analyses van predictieve en consequentiële validiteit, en cesuurbepaling, wat de betrouwbaarheid van die analyses zou verbeteren. Bovendien kan een toets nooit volledig gevalideerd worden (Kane, 2013); het is dus zinloos om daarop te wachten. Nog volgens Kane (2013) moet niet de toets zelf, maar het gebruik ervan worden gevalideerd. Daarom geldt: als de ijkingsstoets bedoeld is om verplicht te worden afgenomen, dan moet deze eerst verplicht worden gemaakt en vervolgens gevalideerd worden voor dat specifieke gebruik, niet vice versa.

Conclusie

Dit onderzoek had als doel de ijkingsstoetsen voor de bacheloropleidingen in IW en CBBGG te valideren én te verbeteren. Vanuit de argumentatiebenadering op validiteit (Kane, 2013) werden de nodige inferenties expliciet gemaakt, gestructureerd en onderzocht. De focus lag daarbij op de voorspellende waarde van de toetsen en op de manier waarop studenten hun deelname ervoeren en er gevolg aan gaven.

Het onderzoek resulteerde in drie concrete aanpassingen in deze ijkingsstoetsen: (i) de splitsing van het wiskundegedeelte in basis- en standaardniveau, (ii) het schrappen van tekstbegripsvragen, en (iii) het toevoegen van wiskundevragen in studierichtingsspecifieke contexten.

Algemeen kunnen volgende vijf adviezen rond beleid en implementatie van ijkings- en soortgelijke toetsen worden gegeven:

- Evalueer kritisch de samenstelling van de toets. Om risicostudenten te identificeren kan het nuttig zijn eenvoudige basiswiskundevragen mee op te nemen, die niet per se het gewenste startniveau weerspiegelen.
- Gebruik weloverwogen cesuren voor het geven van feedback. Overweeg normreferentie om deze te bekomen. De normscore kan terug worden omgezet naar een criteriumscore om te communiceren.
- Identificeer bij de validering zoveel mogelijk positieve en negatieve consequenties. Kies doordacht dewelke vervolgens te kwantificeren. Maak aanpassingen om de negatieve consequenties te vermijden en de positieve te versterken. Zo is een mogelijk gevolg van ijkingsstoetsdeelname dat het beeld dat de deelnemer heeft van de opleiding wordt bijgesteld. Als dat beeld correcter wordt, is dit een positief gevolg, in het andere geval negatief.
- Communiceer duidelijk welke voorbereiding er verwacht wordt van deelnemers.
- Indien een toets met verplichte deelname gevalideerd moet worden, dient deze eerst verplicht te worden alvorens er degelijk gevalideerd kan worden.

De uiteindelijke bedoeling van ijkingsstoetsen is studenten ondersteunen bij hun studiekeuze, een beslissing met vaak levensbepalende gevolgen. De verantwoordelijkheid

voor die keuze ligt bij de student zelf. De verantwoordelijkheid van toetsontwikkelaars ligt in het ontwikkelen van toetsen die optimaal geschikt zijn voor dit doel. Zorgvuldige en empirisch onderbouwde validatie moet daarbij steeds richtinggevend zijn.

Literatuur

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike, (Ed.), *Educational measurement* (2nd ed., pp. 508–600). Washington, DC: American Council on Education.
- Angoff, W.H. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 9–13). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L.J. (1989). Construct validation after thirty years. In R.E. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 147–171). Urbana: University of Illinois Press.
- Cureton, E.E. (1951). Validity. In E.F. Lindquist (Ed.), *Educational measurement* (pp. 621–694). Washington, DC: American Council on Education.
- De Corte, W, Lievens, F., & Sackett, P.R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*, 92, 1380–1393. <https://doi.org/10.1037/0021-9010.92.5.1380>
- Demulder, L., Verschuere, K., Lacante, M., Donche, V. (2023). Identifying exploration profiles for higher education and their relationship with different student variables and outcomes. *European Journal of Psychology of Education*, 38, 1145–1163. <https://doi.org/10.1007/s10212-022-00641-w>
- Fonteyne, L., Acke, S., Hanssens, J., Marconato, A., Schütz, J., Tambuyzer, B., De Laet, T., & Melis, I. (2023). *Validiteitsrapport ijkingstoetsen 2022*.
- Fonteyne, L., Acke, S., Marconato, A., Tambuyzer, B., De Laet, T., & Melis, I. (2022). *Validiteitsrapport ijkingstoetsen 2021*.
- Germeijs, V., & Verschuere, K. (2006). High school students' career decision-making process: A longitudinal study of one choice. *Journal of Vocational Behavior*, 68(2), 189–204. <https://doi.org/10.1016/j.jvb.2005.08.004>
- Germeijs, V., & Verschuere, K. (2007). High school students' career decision-making process: Consequences for choice implementation in higher education. *Journal of Vocational Behavior*, 70, 223–241. <https://doi.org/10.1016/j.jvb.2006.10.004>
- Hallia, M., Kulmala, P., Pursiainen, J., & Nieminen, P. (2025). The change in entrance exam requirements for medical school: impact on prior performance, entrance exam success, and study achievement. *Education Sciences*, 15, 683. <https://doi.org/10.3390/educsci15060683>
- Hanssens, J. (2024a). *Validation of positioning tests at the start of higher STEM education – An argument based approach* [Doctoraal proefschrift, KU Leuven].
- Hanssens, J., Langie, G., & Van Soom, C. (2023a). Students' perceptions of low stakes positioning tests at the start of higher STEM education: A mixed methods approach. *International Journal*

- of Education in Mathematics, Science, and Technology (IJEMST)*, 11(5), 1094–1112. <https://doi.org/10.46328/ijemst.2889>
- Hanssens, J., Van Soom, C., & Langie, G. (2023b). The stability of pre-enrolment prediction of academic achievement: criterion-referencing versus norm-referencing. *Proceedings of the SEFI annual conference 2023*. <https://doi.org/10.21427/NH2E-WS39>
- Hanssens, J., Van Soom, C., & Langie, G. (2024b). Pre-enrollment tests to predict first-year academic achievement in open-admission higher STEM education. *European Journal of Psychology of Education*, 40, 23. <https://doi.org/10.1007/s10212-024-00918-2>
- Hanssens, J., Van Soom, C., De Cock, M., & Langie, G. (2026). Mathematical word problems in positioning tests for STEM students: difficulty, prediction of academic achievement and student perception. [Manuscript ingediend voor publicatie].
- Husbands, A., Rodgeron, M.J., Dowell, J., & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC Medical Education*, 15(1), 1–9. <https://doi.org/10.1186/s12909-015-0424-0>
- Hernandez-Martinez, P., Williams, J., Black, L., Davis, P., Pampaka, M., & Wake, G. (2011). Students' views on their transition from school to college mathematics: rethinking 'transition' as an issue of identity. *Research in Mathematics Education*, 13(2), 199–130. doi.org/10.1080/14794802.2011.585824
- Ijkingstoets.be (2025).
- Kane, M.T. (2012). All Validity is Construct Validity. Or is it? *Measurement: Interdisciplinary Research and Perspectives*, 10(1–2), 66–70. <https://doi.org/10.1080/15366367.2012.681977>
- Kane, M.T. (2013). Validating the Interpretations and Uses of Test Scores, *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635–694. <https://doi.org/10.2466/pro.1957.3.3.635>
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 103). New York, NY: American Council on Education and Macmillan.
- Organization for Economic Cooperation and Development (OECD). (2015). *Universal basic skills: What countries stand to gain*, Paris: OECD Publishing.
- Plank, S.B., & Jordan, W.J. (2001). Effects of Information, Guidance, and Actions on Postsecondary Destinations: A Study of Talent Loss. *American Educational Research Journal*, 38(4), 947–979. <https://doi.org/10.3102/00028312038004947>
- Rigby, C. (2017). Exploring students' perceptions and experiences of the transition between GCSE and AS Level mathematics, *Research Papers in Education*, 32(4), 501–517. <https://doi.org/10.1080/02671522.2017.1318806>
- Sackett, P., De Corte, W., & Lievens, F. (2010). Decision aids for addressing the validity adverse impact trade-off. In J. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 453–472). New York, NY: Routledge.
- Schelfout, S., Wille, B., Fonteyne, L., Roels, E., Deros, E., De Fruyt, F., & Duyck, W. (2022). How accurately do program-specific basic skills predict success in open access higher education? *International Journal of Educational Research*, 111, 101907. <https://doi.org/10.1016/j.ijer.2021.101907>

- Sireci, S.G. (1998). The Construct of Content Validity. *Social Indicators Research*, 45, 83–117. <https://doi.org/10.1023/A:1006985528729>
- Smither, J.W., Reilly, R.R., Millsap, R.E., Pearlman, K., & Stoffey, R.W. (1993). APPLICANT REACTIONS TO SELECTION PROCEDURES. *Personnel Psychology*, 46, 49–76. Tett, L., Cree, V.E., & Chirstie, H. (2017). From further to higher education: transition as an ongoing process. *Higher Education*, 73, 389–406. <https://doi.org/10.1111/j.1744-6570.1993.tb00867.x>
- Toulmin, S.E. (1958). The uses of argument. Cambridge, UK: Cambridge University Press.
- Vandewalle, J.P.L., & Callens, R. (2013). A positioning test mathematics in Flanders for potential academic engineering students. *Proceedings of the 41st SEFI conference*.
- Vlaamse overheid. (2025a). Het Hogeronderwijsregister. <https://www.hogeronderwijsregister.be/>
- Vlaamse overheid. (2025b). *Studieduur in het hoger onderwijs*. Statistiek Vlaanderen. <https://www.vlaanderen.be/statistiek-vlaanderen/onderwijs-en-vorming/studieduur-in-het-hoger-onderwijs>
- Vlaamse overheid. (2025c). Vlaamse toetsen. <https://www.vlaanderen.be/onderwijs-en-vorming/vlaamse-toetsen>

Valid positioning tests in Flanders

Abstract In his dissertation Validation of positioning tests for higher STEM education – An argument-based approach, Jolan Hanssens (2024a) investigated the validity of two Flemish positioning tests. These tests aim to support students in their study choice by providing feedback on their starting competencies. Using an argument-based approach to test validity, the extent to which the tests measure what they intend and to which they can be used to support students was examined.

Based on this research, three modifications to the test composition were implemented. Firstly, the reading comprehension part was removed because it did not contribute to predictive validity and had a negative impact in participants' perception on the test. Secondly, the mathematics part was divided into standard and basic mathematics, the latter easier items being able to identify at-risk students. Thirdly, another mathematics part set in programme-specific contexts was added to underscore the relevance of the test and thereby strengthen its voluntary consequences.

This reflection concludes with five recommendations for the policy and implementation of related tests concerning (i) test composition, (ii) cut-off determination, (iii) weighing consequences, (iv) communication on preparation, and (v) validation with mandatory participation.

Keywords positioning test, test validity, cut-off scores, study success, study choice