

Selection for Medical School

Citation for published version (APA):

Schreurs, S. (2020). *Selection for Medical School: the quest for validity*. Ipskamp Printing BV.
<https://doi.org/10.26481/dis.20200320ss>

Document status and date:

Published: 01/01/2020

DOI:

[10.26481/dis.20200320ss](https://doi.org/10.26481/dis.20200320ss)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

Selection for medical school is an important and somewhat controversial topic. In **Chapter 1**, we explained that especially in the Netherlands, discussions about whether we should be selecting our medical students are flaring up. The biggest argument used to banish selection procedures is that they are expensive and unpredictable. However, every university does it in their own manner, because of which each university has different expenses and predictive values. Currently, it is hard to see the wood for the trees in the forest that is selection: what tools should be used, should they combine cognitive and (inter)personal competencies, what specific competencies or skills should be measured, should we focus on selecting students or doctors?

Current best practice for selection is starting with a job analysis and using the competencies emerging from this job analysis to base your selection procedure on. At Maastricht University, we did this by applying '*constructive alignment*': the outcome frameworks used to assess whether a student is ready to be a doctor by the end of medical school, are also the focus of the clinical and pre-clinical phase of medical school, as well as our selection procedure. In other words, we used backward chaining to reason back from the end goals of medical school to input for our selection procedure. The competencies described in the outcome framework were translated into derived competencies adapted to the level of medical school applicants.

In the empirical studies in this thesis, we took on the quest for validity of this new setup of selection for medical school. To be able to assess the usefulness of this '*constructive alignment*' approach to selection, we applied a '*modern validity framework*'. Modern validity frameworks explain all validity as construct validity (i.e. are you measuring what you intend to measure), and state that validity is a continuum and an ongoing process related to inferences, not tools or assessments in themselves. The validity framework we derived from multiple modern validity frameworks focused on supporting construct validity using five sources of evidence: content, response processes, internal structure, relation to other variables, and consequences. This validity framework was applied to the Maastricht selection procedure over the course of four empirical studies.

In the empirical study described in **Chapter 2**, the focus was on evidence for validity based on the '*relation to other variables*' pillar of the validity framework: the students' performance throughout the medical bachelor. The aim of this study was to examine if the outcome-based selection procedure developed at Maastricht University was predictive of study success in the bachelor phase. To do so, the relationship between performance at selection and later study success across the three-year bachelor program was examined in three student cohorts (2011, 2012 and 2013). These cohorts were chosen because in these years, there were two parallel admission routes providing entry to the medical bachelor curriculum: the decentralized (i.e. individual to each university) selection procedure or the national weighted lottery. In this study, two groups were compared. Firstly, the students who were selected in the selection

procedure and therefore got into the medical curriculum through selection (selection-positive students). Secondly, the students who entered the medical bachelor curriculum through the national weighted lottery procedure after having been rejected in the selection procedure (selection-negative students). Study results were compared between these selection-positive and selection-negative students.

The student results taken into account in this study were all graded assessments in the medical bachelor curriculum, divided into four categories: cognitive, (inter)personal, mixed and general. It was found that selection-positive students outperformed their selection-negative counterparts throughout the entire bachelor program on assessments measuring cognitive (e.g. written exams), (inter)personal (e.g. communication and reflection skills) and combined outcomes (i.e. objective structured clinical examinations). No significant effects on general parameters (e.g., drop-out and delay) were found. Of the 30 outcome variables, selection-positive students scored significantly higher in 11 cases. Fifteen other, non-significant group differences were also in favor of the selection-positives. An overall comparison using a sign test indicated a significant difference between both groups ($p < 0.001$), despite equal pre-university Grade Point Averages (GPAs).

All in all, it seemed that the use of a constructively aligned selection approach addresses some of the predictive validity limitations of commonly-used selection tools: a constructively aligned selection procedure appears to be capable of predicting multiple categories of outcomes in the medical bachelor. Selection-positive students significantly outperformed their selection-negative counterparts across a range of cognitive, (inter)personal, and mixed outcomes throughout the entire three-year bachelor curriculum.

In **Chapter 3**, a follow-up study is presented. In this chapter, we investigated whether the selection procedure explained above was predictive of the students' performance in the clinical master phase of the medical curriculum as well. Thus, in terms of the validity framework, it also studied the selection procedure's relation to other variables. The aim of this study was to assess the relationship between performance at selection and during the clinical years of a medical program. Again, the context was a medical curriculum in which the selection procedure, curriculum content and assessments are all aligned with the same outcome framework (i.e. the Canadian Medical Education Directives for Specialists; CanMEDS). The set-up of this study was similar to the study described in Chapter 2. Two groups of students were compared: those selected via an outcome-based selection procedure (selection-positive students) versus those rejected in this procedure who entered the program through a national, GPA-based lottery procedure (selection-negative students).

Performances of both groups on all seven CanMEDS-roles were compared during clinical rotations, for each of the three master years. Students' performances on all competences were recorded in an electronic portfolio, in a programmatic manner: students gathered quantitative information (e.g., on knowledge tests) as well as qualitative feedback from different assessors (e.g., medical specialists, peers, nurses,

patients) at different times for different assignments. At the end of a year, the exam committee appraised all information in the portfolio and provided a final judgement for each competency: below expectations, as expected, or exceeds expectations. Data were compared for the three cohorts starting their medical bachelor curriculum in 2011, 2012 or 2013, since these are the only cohorts in which both selection-positive and selection-negative students were admitted to the medical curriculum. It was found that selection-positive students significantly outperformed the selection-negative students in all three master years, and that the differences between the two groups increased over time. In the first year, differences were apparent in the roles of Communicator, Collaborator and Professional, in the second year also for the roles of Organizer and Health Advocate, and in the third year in the role of Academic as well.

The conclusion of this study was that the constructively aligned selection procedure had an increasing predictive value across the clinical years of a medical curriculum. This suggests that constructive alignment of selection, curriculum and assessment to ultimate outcomes is effective in creating a selection procedure predictive of clinical performance.

In **Chapter 4**, we moved away from evidence for validity based on relation to other variables, and went on to investigate the content and internal structure of our selection procedure more closely. For some time now, medical school selection finds itself in the paradoxical situation in which selection tools may predict study outcomes, but which constructs are actually doing the predicting is unknown (the ‘black box of selection’). Therefore, this study focused on those constructs, answering the question: do the internal structures of the tests in a constructively aligned selection procedure reflect the content that was intended to be measured? To do so, we had to first explicate what the content of the selection procedure was supposed to be, and then delve deep into the psychometric qualities of the tests in the procedure to see whether the psychometric qualities were a fit with the content we were intending to measure.

First, we examined content-related evidence pertaining to the creation and application of the competency-based selection blueprint. Constructive alignment was achieved by finding a well-established job analysis (i.e. the CanMEDS) and translating the competencies in the CanMEDS to the level of an applicant in the selection procedure. To do so, ‘backward chaining’ was applied. The resulting selection procedure was a multi-tool, CanMEDS-based procedure consisting of two rounds. The focus of this study was on the second round, which comprised a Video-based Situational Judgement Test (focused on (inter)personal competencies) and a Written Aptitude Test (reflecting a broader array of CanMEDS competencies). We found that the set-up of the selection procedure was a robust, transparent and replicable process.

Second, the internal structure of the selection tests was investigated by connecting applicants’ performance on these tests to the predetermined blueprint using Cognitive Diagnostic Modeling (CDM). CDM is capable of finding latent variables when there is

multidimensionality in the data, both between and within items, as was the case in our data: different items measured different competencies, but most items also measured multiple competencies. CDM is related to Confirmatory Factor Analysis: the structure (in this case the selection procedure's blueprint) is provided, and CDM looks at whether that structure is indeed found in the data, or whether alterations to the structure make more sense on the basis of the data provided. The data indicated 89% overlap between the expected and measured constructs.

All in all, our results supported the notion that the focus placed on creating the right content and following a competency-blueprint was effective in terms of internal structure: most items measured what they were intended to measure. This way of linking a predetermined blueprint to the applicants' results sheds light into the 'black box of selection' and can be used to support the construct validity of selection procedures.

An underexposed pillar of the validity framework named in the beginning of this summary chapter is the 'consequences' one. The focus of the empirical study in **Chapter 5** was on one specific part of consequences: cost-efficiency, specifically from the perspective of the medical school. This research is especially important, as resources for medical education are becoming more constrained, whereas accountability in medical education is increasing. In this constrictive environment, medical schools need to consider and justify their selection procedures in terms of costs and benefits. To date, however, there have been no studies focusing on this aspect of selection. Therefore, in this study, we aimed to examine and compare the costs and benefits of two different approaches to admission into medical school: a constructively aligned, multimethod selection process versus a GPA-weighted lottery procedure. Our goal was to assess the relative effectiveness of each approach and to compare these in terms of benefits and costs from the perspective of the medical school.

The study was conducted in the medical curriculum at Maastricht University, where a decentralized selection process and a weighted lottery procedure ran in parallel for three years (2011–2013). The costs and benefits of the selection process were compared with those of the lottery procedure over three student cohorts throughout the bachelor program. The extra costs of selection represented the monetary investment of the Institute for Education that organized the medical curriculum in conducting the selection procedure; the benefits were derived from the increase in income generated by the prevention of dropout and the reductions in extra costs due to decreases in the repetition of blocks and objective structured clinical examinations. We found that the constructively aligned selection procedure cost about €139,000 for a full cohort of students ($n = 286$). The lottery procedure came with negligible costs. However, the average benefits of selection compared with the lottery system added up to almost €207,000 per cohort of students. This resulted in an overall benefit of selection over lottery to the tune of €68,000.

In conclusion, this study not only showed that conducting a cost–benefit comparison is feasible in the context of selection for medical school, but also that an ‘expensive’ selection process can be cost beneficial in comparison with an ‘inexpensive’ lottery system.

Chapter 6 is a commentary to a meta-analysis by Jon Foo and colleagues, in which they showed that studies on costs are extremely scarce in Health Professions Education (HPE), and studies including value are even more rare. Furthermore, the quality of most of the available research appeared to be substandard and had not increased since the beginning of this century. These findings mirror the practice of higher education institutes: decisions on which educational activities to invest in are often not fueled by full empirical economic evaluations. We indicated that one educational area in which cost evaluations are particularly interesting is that of selection. While medical schools are under increasing pressure to justify their unique and often expensive selection procedures in terms of costs and value, research on the costs and especially the value of selection procedures, or even the tools within these procedures, is sorely lacking.

Importantly, high quality cost-evaluation research is not only relevant at the level of individual educational institutes, but also on a national level. As an example, we referred to the situation in the Netherlands: in 2017, medical schools were obliged to switch from a national weighted lottery to decentralized selection procedures which differ from school to school. Because several of these procedures were shown to be insufficiently predictive to warrant the monetary investments, the cost-effectiveness of selection in general was questioned and reintroduction of a national lottery system is proposed. This dispute has even sparked discussion at the government level. In our opinion, research into the general construct validity of the selection procedures, including high-quality full economic evaluations, are crucial to support an evidence-based governmental decision on this issue.

All in all, there is a lack of high-quality economic evaluations in HPE in general and selection in particular. This puts the stakeholders in a precarious position in which they have to remain accountable regarding their limited financial resources while choosing between educational activities for which the costs are unknown. Therefore, educational researchers should join efforts to conduct more economic evaluations to provide relevant stakeholders with economically valid arguments to make well-informed decisions that benefit the quality of education.

In **Chapter 7**, the General Discussion, the main findings of all previous chapters were summarized and implications and suggestions for the practice of and research on selection were debated. Overall, the empirical studies included in this dissertation all add information to a validity argument for a selection procedure based on the principles of constructive alignment. All of the studies can be integrated into the validity framework established in Chapter 1 of this dissertation:

- ✓ *Content*: a thorough job analysis on which a test blueprint is based results in a robust, transparent and replicable process

- ✓ *Response processes*: first, preliminary data suggest no effect of changes in weighting of portfolio subparts on age and gender of selected students
- ✓ *Internal structure*: about 90% overlap between the intended and actually measured competencies, fair accuracies per competency, high intra- and inter-rater reliability
- ✓ *Relation to other variables*: predictive for pre-clinical performance during a medical bachelor curriculum (cognitive, [inter]personal and mixed results), increasingly predictive for day-to-day clinical performance during the master phase
- ✓ *Consequences*: ‘expensive’ selection appears to be more cost-effective than a ‘cheap’ GPA-weighted lottery

These findings can have important repercussions for several (inter)national discussions: the cognitive versus (inter)personal discussion should not so much be a black-and-white choice, but more of a continuum, focusing on measuring relevant competencies along the entirety of this continuum. Furthermore, blindly relying on evidence of tools and disregard for the content causes mixed results in research as well as practice. Also, applying a job analysis to determine what to measure in a selection procedure is paramount. Importantly, there currently is a discussion on whether selection is worth the hassle. Results differ considerably between universities, but the focus in the discussion is on the negative results.

With this dissertation we hope to stop that tilting and direct more attention towards the careful planning of constructs and thorough evaluation of selection procedures, to achieve fair and predictive selection procedures. The implications of this thesis for theory, research and practice are manifold: the use job analyses as the starting point for a selection procedure; the use of innovative methods for selection research (e.g. Cognitive Diagnostic Modelling and cost-effectiveness); and the application of modern validity theories.

The main conclusion from the current thesis is that explicit constructive alignment can be seen as a way forward for selection. Furthermore, it is recommended to evaluate selection procedures using modern validity frameworks to achieve a more general understanding of validity. A promising test theory to help in applying those validity theories is Cognitive Diagnostic Modelling. In combination, constructive alignment, modern validity frameworks and innovative research methods (e.g. CDM and cost-effectiveness) could really help the field of selection forward and towards a ‘holy grail’: a predictive, reliable, valid, fair, acceptable and cost-effective selection procedure.

Samenvatting

Selectie voor de geneeskundeopleiding is een belangrijk en enigszins controversieel onderwerp. Hoofdstuk 1 beschrijft de discussies, voornamelijk in Nederland, over of geneeskunde studenten wel of niet geselecteerd zouden moeten worden. Het belangrijkste argument om van selectie af te stappen is dat selectieprocedures duur zijn, maar niet voorspellen. De kanttekening hierbij is dat de selectieprocedures verschillen per universiteit, waardoor iedere universiteit te maken heeft met andere kosten en verschillende voorspellende waardes. Op dit moment is het moeilijk om door de bomen het bos nog te zien op het gebied van selectie: welke methodes/toetsen moeten we gebruiken, moeten we daarin cognitieve en (inter)persoonlijke competenties combineren, welke specifieke competenties of vaardigheden moeten we meten en moeten we ons richten op het selecteren van de beste studenten of juist van de beste toekomstige artsen?

De huidige '*best practice*' binnen selectie is om te beginnen met een analyse van de taken van een toekomstige arts en de competenties die daarvoor nodig zijn. Die competenties worden vervolgens gebruikt als de basis voor het ontwerp van de selectieprocedure. Aan de Universiteit van Maastricht (UM) hebben we dit gedaan door '*constructive alignment*' toe te passen, met het raamplan voor de geneeskundeopleiding als basis. Dit raamplan beschrijft de uitkomsten van de geneeskundeopleiding: de competenties en de niveaus binnen die competenties die de studenten aan het einde van hun opleiding behaald moeten hebben. Het raamplan wordt dus gebruikt om te toetsen of een student klaar is om arts te zijn tegen het einde van de geneeskundeopleiding, maar ook het curriculum en de toetsing in de klinische (master)fase en pre-klinische (bachelor)fase van de UM zijn gebaseerd op deze competenties. Door middel van '*backward chaining*' werden deze competenties ook de focus van de selectieprocedure: de einddoelen van de geneeskundeopleiding zijn terug geredeneerd en aangepast aan het niveau van een kandidaat in de selectieprocedure. De resulterende 'afgeleide competenties' vormden de blauwdruk voor de selectieprocedure.

De empirische studies in dit proefschrift vormen een 'quest' voor validiteit van deze nieuwe opzet van selectie voor de geneeskundeopleiding. Om te onderzoeken hoe nuttig *constructive alignment* is als aanpak voor het opzetten van een selectieprocedure hebben we een zogenaamd modern validiteitsraamwerk toegepast. Moderne validiteitsraamwerken leggen alle validiteit uit als constructvaliditeit (meet dit instrument wat het zou moeten meten?). Binnen deze raamwerken wordt validiteit gezien als een continuüm en is validering een aanhoudend proces dat gerelateerd is aan gevolgtrekkingen (specifiek gebruik van het middel/de toets) en niet aan de middelen of toetsen zelf. Het validiteitsraamwerk dat we gebruiken in dit proefschrift is een afgeleide en combinatie van meerdere moderne validiteitsraamwerken, en is gericht op het ondersteunen van constructvaliditeit door middel van vijf bronnen van bewijs: inhoud, reactieprocessen, interne structuur, relatie met andere variabelen en

consequenties. Dit validiteitsraamwerk werd toegepast op de selectieprocedure voor de geneeskundeopleiding in Maastricht, in vier empirische studies.

Het empirisch onderzoek dat beschreven wordt in **Hoofdstuk 2** was gericht op het verzamelen van validiteitsbewijs op basis van de relatie tussen selectie en andere variabelen: de prestaties van de studenten gedurende de geneeskunde bachelor. Het doel van de studie was om te onderzoeken of de selectieprocedure die ontwikkeld is in Maastricht ook daadwerkelijk voorspellend is voor het studiesucces in de geneeskundebachelor. Om dit doel te kunnen bereiken, hebben we de relatie tussen de prestatie in de selectieprocedure en het latere studiesucces in de drie jaar durende bachelor onderzocht in drie studentencohorten (2011, 2012 en 2013). Deze cohorten werden gekozen, omdat er in deze jaren twee parallelle routes waren om het geneeskundecurriculum binnen te komen: de decentrale (oftewel voor iedere individuele universiteit verschillende) selectieprocedure, of de centrale (nationale) gewogen loting. In dit onderzoek zijn twee groepen vergeleken: de studenten die geselecteerd werden in de lokale selectieprocedure (*selectie-positieve studenten*) en de studenten die afgewezen werden in diezelfde selectieprocedure maar alsnog het geneeskundecurriculum binnenkwamen via de nationale gewogen loting (*selectie-negatieve studenten*). Alle beoordelingen die studenten gedurende het bachelor curriculum kregen, werden meegenomen. Deze beoordelingen werden onderverdeeld in vier categorieën: cognitief, (inter)persoonlijk, een combinatie van deze twee, en algemene gegevens (uitval en studievertraging).

De bevindingen waren als volgt: de selectie-positieve studenten presteerden beter dan hun selectie-negatieve medestudenten gedurende het hele bachelor programma, op zowel cognitieve (bijvoorbeeld geschreven bloктоetsen), (inter)persoonlijke (bijvoorbeeld communicatie- en reflectie-vaardigheden) als gecombineerde toetsen (stationstoetsen). Voor de algemene uitkomstvariabelen werden geen significante effecten gevonden. Van de in totaal 30 uitkomstvariabelen deden de selectie-positieve studenten het significant beter op 11 uitkomsten. Vijftien andere, niet-significante groepsverschillen waren ook in het voordeel van de selectie-positieve studenten. Een algehele vergelijking van alle uitkomsten door middel van een '*sign test*' liet dan ook een significant verschil tussen de beide groepen zien ($p < 0.001$), ondanks de vergelijkbare middelbare school gemiddelden van de beide groepen.

Al met al lijkt een selectieaanpak die gebruik maakt van '*constructive alignment*' het gebrek aan voorspellende waarde dat gevonden wordt bij andere, veelgebruikte selectietoetsen te kunnen opheffen: een selectieprocedure die helemaal in lijn is gebracht met de inhoud, toetsing en de eindtermen van de studie kan uitkomsten in de geneeskundebachelor voorspellen. Selectie-positieve studenten presteren significant beter dan hun selectie-negatieve medestudenten over een reeks van cognitieve, (inter)persoonlijke en gecombineerde uitkomsten gedurende de hele drie jaar durende geneeskunde bachelor.

In **Hoofdstuk 3** wordt een vervolgstudie beschreven. In dit hoofdstuk werd onderzocht of de hierboven beschreven selectieprocedure ook voorspellend was voor de

prestaties van de studenten in de klinische masterfase van het geneeskunde curriculum. Op deze manier draagt ook deze studie bij aan het bewijs voor validiteit op basis van de relatie van de selectieprocedure met andere variabelen. Het doel van dit onderzoek was om de relatie tussen de prestaties in de selectieprocedure en tijdens de klinische jaren van het geneeskunde programma vast te stellen. Belangrijk hierbij is weer de context: de selectieprocedure, de inhoud van het curriculum en de toetsen tijdens de studie zijn allemaal in lijn met hetzelfde uitkomstenraamwerk, namelijk het Raamplan Artsopleiding 2009. De opzet van deze studie lijkt erg op de studie beschreven in hoofdstuk 2: twee groepen studenten werden vergeleken, namelijk de studenten die geselecteerd werden via de selectieprocedure (selectie-positieve studenten) versus de studenten die afgewezen werden in de selectieprocedure en het geneeskunde curriculum alsnog binnenkwamen via een nationale, op basis van middelbare school cijfers gewogen, loting procedure (de selectie-negatieve studenten).

Prestaties van beide groepen studenten op alle zeven in het raamplan beschreven competenties (Medisch deskundige, Communicator, Samenwerker, Organisator, Gezondheidsbevorderaar, Academicus en Beroepsbeoefenaar) werden vergeleken gedurende de klinische rotaties (co-schappen), voor ieder masterjaar. Deze prestaties werden bijgehouden in een elektronisch portfolio. In het portfolio verzamelen studenten zowel kwantitatieve informatie (bijvoorbeeld resultaten van kennistoetsen) als kwalitatieve feedback van verschillende beoordelaars (bijvoorbeeld medisch specialisten, medestudenten, verpleging, patiënten) op verschillende momenten voor verschillende taken. Op deze manier wordt een grote bron aan informatie opgebouwd. Aan het einde van het jaar checkt de examencommissie alle, door de student verzamelde, informatie in het portfolio en wordt een definitief, programmatisch oordeel voor iedere competentie gegeven: beneden verwachting, naar verwachting, of boven verwachting. Data van de drie cohorten die begonnen aan hun geneeskunde bachelor curriculum in 2011, 2012 of 2013 werden vergeleken, aangezien dit de enige cohorten zijn waarin zowel selectie-positieve als selectie-negatieve studenten toegelaten werden. De selectie-positieve studenten bleken significant beter te presteren dan de selectie-negatieve studenten in alle drie masterjaren, en de verschillen tussen deze twee groepen bleken toe te nemen in de tijd. In het eerste jaar deden de selectie-positieve studenten het significant beter in de rollen Communicator, Samenwerker en Beroepsbeoefenaar, in het tweede en derde jaar hiernaast ook nog in de rol van Organisator en Gezondheidsbevorderaar, en in het derde jaar bleek dat hier bovenop ook te gelden voor de rol van Academicus.

De conclusie van deze studie was dat een selectieprocedure die '*constructively aligned*' is met het geneeskunde curriculum en het einddoel van de studie, een toenemende voorspellende waarde heeft in de klinische fase van het geneeskunde curriculum. Dit suggereert dat *constructive alignment* van selectie, curriculum en toetsing met de einddoelen van de opleiding een effectieve manier kan zijn om een selectie 'aan de poort' op te zetten die voorspellend is voor de klinische prestaties van studenten aan het einde van hun geneeskunde studie.

Hoofdstuk 4 richtte zich in meer detail op de inhoud en interne structuur van de selectieprocedure. Al een hele tijd bevindt het veld van selectie voor geneeskunde zich in de paradoxale situatie waarin we weten dat een aantal selectietoetsen bepaalde studie-uitkomsten voorspellen, maar welke constructen (onderwerpen/competenties) binnen de selectieprocedure nou eigenlijk zorgen voor die voorspellende waarde weten we niet (dit noemen we ook wel de *'black box'* in selectie). Daarom was deze studie specifiek gericht op het onderzoeken van welke constructen gemeten moesten worden, hoe die geïntegreerd konden worden in een procedure en in hoeverre de resulterende selectieprocedure erin slaagde om deze constructen daadwerkelijk te meten. Daarmee werd volgende vraag beantwoord: passen de psychometrische kwaliteiten van de toetsen in een *'constructively aligned'* selectieprocedure bij de inhoud die gemeten zou moeten worden? Om dit te kunnen onderzoeken werd de beoogde inhoud van de selectieprocedure uiteengezet, om vervolgens de psychometrische kwaliteiten te achterhalen en te onderzoeken of deze passen bij die beoogde inhoud.

Als eerste hebben we naar het inhoud-gerelateerde bewijs gekeken, met specifieke focus op de manier van opzetten en toepassen van de competentie-gebaseerde selectieblauwdruk. Zoals eerder benoemd werd *constructive alignment* bereikt door de eindtermen van de geneeskundeopleiding te raadplegen en te vertalen naar het niveau van een deelnemer aan de selectieprocedure. Hiervoor werd *backward chaining* toegepast. De resulterende selectieprocedure omvatte meerdere toetsen in twee rondes, waarvan de inhoud steeds gebaseerd was op de eindtermen. De focus van dit onderzoek was de tweede ronde, waarin een video-gebaseerde situationele beoordelingstoets (een Situational Judgement Test, gericht op het meten van (inter)persoonlijke competenties) en een schriftelijke aanlegtoets (een Aptitude Test, waarin een breder spectrum van de eindtermen competenties aan bod komt) werden afgenomen. De opzet van de selectieprocedure bleek robuust en transparant, en het proces van het ontwikkelen van de inhoud was reproduceerbaar.

Als tweede werden de interne structuren van de selectietoetsen onderzocht door de prestaties van de deelnemers op de selectietoetsen te relateren aan de vooraf opgestelde blauwdruk van de toetsen (deze beschrijft welke items in de toetsen bedoeld zijn om welke competenties te meten) door middel van *'Cognitive Diagnostic Modelling'* (CDM). CDM is ertoe in staat latente variabelen te vinden, zelfs als er sprake is van multidimensionaliteit in de data, zowel binnen als tussen items. Dit was in onze data het geval: verschillende items meten verschillende competenties, maar de meeste items meten ook meerdere items. CDM is gerelateerd aan *'Confirmatory Factor Analysis'*: de structuur (in dit geval de blauwdruk van de selectietoetsen) wordt ingevoerd in de analyse, en CDM bepaalt of diezelfde structuur inderdaad terug te vinden is in de data, of dat er aanpassingen nodig zijn in de structuur zodat deze beter past bij de data. Uit deze analyse bleek dat er sprake was van een overlap van 89% tussen de beoogde en daadwerkelijk gemeten constructen (competenties).

Alles bij elkaar genomen ondersteunen deze resultaten het idee dat de focus die we geplaatst hebben op het creëren van de juiste inhoud en het volgen van een

competentie-blauwdruk effectief was in termen van de interne structuur van de selectietoetsen: het overgrote deel van de items meet inderdaad wat ze zouden moeten meten. Deze manier van linken van een vooraf opgestelde blauwdruk aan de daadwerkelijke resultaten van de deelnemers werpt licht in de 'black box' van selectie, en ondersteunt tevens de constructvaliditeit van de selectieprocedures.

Een andere onderbelichte pilaar binnen het validiteitsraamwerk dat in het begin van deze samenvatting genoemd werd, zijn de consequenties. Daarom lag de focus van het empirische onderzoek in **Hoofdstuk 5** op een specifiek onderdeel van deze consequenties: de kosteneffectiviteit van selectie vanuit het oogpunt van de geneeskundeopleiding. Dit onderzoek is belangrijk omdat de middelen voor medisch onderwijs meer en meer beperkt worden, terwijl er tegelijkertijd steeds meer verantwoording over de besteding van deze middelen afgelegd moet worden. In deze beperkende omgeving moeten de geneeskundeopleidingen nadenken over hun selectieprocedures en moeten deze verdedigbaar zijn in termen van kosten en baten. Tot op heden zijn er echter nog geen studies die zich op dit aspect van selectie richten. Daarom was het doel van deze studie om de kosten en opbrengsten van twee verschillende benaderingen van toelating tot het geneeskundecurriculum te vergelijken: een '*constructively aligned*' selectieprocedure met meerdere toetsen versus een lotingsprocedure die gewogen is naar middelbare school gemiddelden. Ons doel was om de relatieve effectiviteit van de beide aanpakken te onderzoeken en te vergelijken in termen van kosten en opbrengsten vanuit het perspectief van de geneeskundeopleiding.

Deze studie is gedaan binnen het geneeskundecurriculum aan de UM, waar in de periode van 2011 tot 2013 een decentrale selectieprocedure en gewogen loting procedure parallel toegepast werden. De kosten en opbrengsten (over de gehele bachelor-periode) van de selectieprocedure werden vergeleken met die van de lotingprocedure voor drie studentengroepen (2011, 2012, en 2013).

De extra kosten voor selectie bestonden uit de financiële investering van de geneeskundeopleiding in het opzetten en uitvoeren van de selectie; de 'opbrengsten' werden afgeleid van de toename in inkomsten door uitval te voorkomen (voor een student die uitvalt in jaar 1 krijgt de geneeskundeopleiding 2 jaar geen inkomsten, voor uitval in jaar 2 krijgt de opleiding 1 jaar geen inkomsten) en de extra kosten door herhaling van blokken en stationstoetsen te verminderen. De organisatie van een '*constructively aligned*' selectieprocedure kostte ongeveer €139.000 voor een compleet cohort van 286 studenten. De kosten van de lotingsprocedure waren voor de geneeskundeopleiding verwaarloosbaar (deze kosten werden nationaal gedragen). De gemiddelde 'opbrengsten' van de selectieprocedure ten opzichte van de lotingprocedure bedroegen bijna €207.000 per cohort studenten. Dit resulteert in een voordelige balans van selectie ten opzichte van loting van ongeveer €68,000.

Concluderend laat deze studie niet alleen zien dat het mogelijk is om kosteneffectiviteitsstudies te doen binnen de context van selectie voor de

geneeskundeopleiding, maar ook dat een 'dure' selectieprocedure kosteneffectief kan zijn ten opzichte van een 'goedkoop' lotingsstelsel.

Hoofdstuk 6 betreft een commentaar op een meta-analyse van Jonathan Foo en collega's. Zij tonen aan dat studies gericht op kosten heel zeldzaam zijn binnen het 'Health Professions Education' veld, en dat studies waarin ook de waarde (baten, opbrengsten) worden meegenomen nog minder voorkomen. Verder bleek de kwaliteit van de meerderheid van de beschikbare literatuur niet aan de maat te zijn, en bleek die kwaliteit ook niet toegenomen te zijn in de afgelopen decennia. Deze bevindingen spiegelen de situatie in de praktijk van het hoger onderwijs: beslissingen over investeringen in onderwijsactiviteiten worden vaak niet gedreven door volledige empirische en economische evaluaties. Een specifiek veld binnen het onderwijs waarin kostenevaluaties bijzonder interessant zijn is de selectie: geneeskundeopleidingen moeten hun unieke en vaak dure selectieprocedures verdedigen in termen van kosten en baten, terwijl onderzoek naar de kosten en vooral de baten van de selectieprocedures, of zelfs maar de toetsen binnen deze procedures, uitermate zeldzaam is.

Belangrijk is ook dat goede kosten-baten evaluaties niet alleen belangrijk zijn voor individuele instituten, maar ook op nationaal niveau. In Nederland, bijvoorbeeld, zijn alle geneeskundeopleidingen sinds 2017 verplicht om hun studenten te selecteren. De nationale gewogen loting werd afgeschaft, en iedere universiteit moest een eigen selectieprocedure ontwikkelen, die per universiteit verschillend is. Aangezien de voorspellende waarde van deze procedures wisselend en vaak beperkt bleek te zijn, werd de kosteneffectiviteit van selectie in het algemeen in twijfel getrokken en is er discussie ontstaan over eventuele herintroductie van de gewogen loting. Deze discussie bereikt zelfs de regering. Vanuit ons oogpunt zijn onderzoeken naar de algemene constructiviteit van selectieprocedures (in plaats van alleen onderzoek naar de voorspellende waarde), inclusief volledige economische evaluaties van hoge kwaliteit, cruciaal voor het ondersteunen van een parlementaire beslissing op dit punt.

Al met al is er dus een groot tekort aan economische evaluaties van hoge kwaliteit in het veld van het hoger onderwijs in het algemeen, en binnen het gebied van selectie in het bijzonder. Dit is problematisch voor de belanghebbenden, omdat zij aansprakelijk zijn voor het gebruik van de gelimiteerde financiële middelen, terwijl ze moeten kiezen tussen onderwijsactiviteiten zonder prijskaartjes. Meer samenwerking tussen onderwijskundige onderzoekers op het gebied van economische evaluaties is van belang om belangrijke beslissingen te kunnen baseren op economisch valide argumenten, hetgeen de kwaliteit van onderwijs ten goede zal komen.

In **Hoofdstuk 7**, de Algemene Discussie, zijn alle hoofdbevindingen uit de eerdere hoofdstukken samengevat en werden implicaties en suggesties voor de praktijk en het onderzoek rondom selectie bediscussieerd. Overkoepelend kunnen we zeggen dat de empirische studies in dit proefschrift bijdragen aan een validiteitsargument voor een selectieprocedure die gebaseerd is op het principe van '*constructive alignment*'. Alle

studies kunnen geïntegreerd worden in het validiteitsraamwerk dat in Hoofdstuk 1 uiteen werd gezet:

- ✓ *Inhoud*: een grondige analyse van welke competenties er nodig zijn, waarop een blauwdruk voor de toetsen gebaseerd wordt, resulteert in een robuust, transparant en herhaalbaar proces.
- ✓ *Reactieprocessen*: eerste, preliminaire data laat zien dat het variëren met wegingen in het portfolio geen effect heeft op leeftijd en geslacht van de geselecteerde studenten
- ✓ *Interne structuur*: er is ongeveer 90% overlap tussen de beoogde en daadwerkelijk gemeten competenties, met redelijke betrouwbaarheden per competentie en hoge intra- en inter-beoordelaar betrouwbaarheid
- ✓ *Relatie met andere variabelen*: de selectieprocedure is voorspellend voor preklinische prestaties tijdens de geneeskundebachelor (voor cognitieve, (inter)persoonlijke en gecombineerde uitkomsten) en is in toenemende mate voorspellend voor de dagelijkse klinische prestaties tijdens de masterfase
- ✓ *Consequenties*: een 'dure' selectieprocedure is uiteindelijk kosten-effectiever dan een 'goedkope' lotingprocedure op basis van middelbare school gemiddelden

Deze bevindingen hebben belangrijke repercussies voor een aantal (inter)nationale discussies: de cognitieve versus (inter)persoonlijke competenties-discussie zou geen zwart-wit keuze moeten zijn, maar selectie moet gefocust zijn op het meten van relevante competenties over het hele continuüm. Verder zorgen het blind vertrouwen op bewijs voor individuele toetsen/middelen en een gebrek aan focus op en validering van de inhoud voor inconsistente resultaten in onderzoek en in de praktijk. Het uitvoeren van een analyse welke competenties er nodig zijn als geneeskundestudent en dokter moet voorop staan bij het vaststellen van de inhoud die in de selectieprocedure gemeten moet worden. Ook de discussie over of de opbrengst van selectie eigenlijk wel de moeite die het kost waard is, is van belang: de resultaten verschillen behoorlijk tussen universiteiten, maar de focus in de discussie wordt vaak gelegd op de negatieve resultaten. Met dit proefschrift hopen we die neiging om vooral op negatieve resultaten te focussen te stoppen, zodat we niet het kind met het badwater weggooien. Er zou meer aandacht gevestigd moeten worden op het zorgvuldig plannen van de constructen en het grondig evalueren van de procedures, om zo te komen tot goed onderbouwde, kosteneffectieve en voorspellende selectieprocedures.

De implicaties van dit proefschrift voor theorie, onderzoek en praktijk zijn veelvoudig: het beginnen met het opzetten van een procedure om te analyseren welke competenties er nodig zijn in de studie en latere klinische praktijk; het gebruik van innovatieve methoden van onderzoek (bijvoorbeeld *Cognitive Diagnostic Modelling* en kostenevaluaties); en de toepassing van moderne validiteitstheorieën.

De voornaamste conclusie uit dit proefschrift is dat expliciete '*constructive alignment*' een goede weg vooruit is voor selectie. Daarnaast is het aanbevelenswaardig om

selectieprocedures te evalueren door middel van moderne validiteitsraamwerken, om zo een breder begrip van de validiteit van de procedures te verkrijgen. Een veelbelovende testtheorie die hieraan zou kunnen bijdragen is *Cognitive Diagnostic Modelling*. Een combinatie van deze drie aanbevelingen, '*constructive alignment*', moderne validiteitsraamwerken en innovatieve methoden (waaronder CDM), zou het veld van selectie kunnen helpen dichterbij die 'heilige graal' te komen: een voorspellende, betrouwbare, valide, eerlijke, acceptabele en kosteneffectieve selectieprocedure.