

Feedback in the context of high-stakes assessment

Citation for published version (APA):

Harrison, C. J. (2017). *Feedback in the context of high-stakes assessment: can summative be formative?*. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20170920ch>

Document status and date:

Published: 01/01/2017

DOI:

[10.26481/dis.20170920ch](https://doi.org/10.26481/dis.20170920ch)

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

**Feedback in the context
of high-stakes assessment:
Can summative be formative?**

Summary

Feedback can be an extremely powerful phenomenon, one of the most powerful ways to stimulate performance improvement. Unfortunately, it frequently fails to reach its potential. Reasons for this include poor delivery of feedback, problems with the reception of feedback, or because learners are unable to translate the feedback into actual learning activities that would lead to positive change. The role of feedback within the context of summative assessment has previously received little attention. This thesis therefore set out to consider the challenges of receiving feedback within this context, and in particular to challenge the common assumption that assessments have either a formative or a summative purpose and that it is difficult for a single assessment to serve both purposes at the same time.

There are several reasons why it would seem sensible to combine the formative and summative purposes of assessment. High-stakes assessments typically create large amounts of data regarding a student's strengths and weaknesses, but this information is reduced down to a binary pass-fail decision on progression to the next stage. While this is an understandable approach, it has the potential to mislead as much rich information about a learner has been lost in the process, and fails to prepare students for independent self-regulated learning. Detailed information from assessments has the potential to support students to identify their own strengths and weaknesses. Practical assessments in medical education are extremely expensive, so it also makes financial sense to maximise the value which can be obtained from them.

On the other hand the delivery of feedback in these circumstances presents practical challenges, with a need for significant faculty resources, although these can be potentially overcome, in part by the use of information technology. However, learners' receptivity to the feedback is more challenging; it is uncertain whether they will engage with the feedback in a meaningful way in this context. There are potential barriers at both the individual and cultural level to the reception of feedback, but these have not been studied in the context of summative assessment. Summative assessment is known to have adverse effects on other aspects of the learning process, so it is plausible that it could have undesirable effects on feedback receptivity. Alternatives to the summative assessment culture towards one based on an assessment for learning have been proposed, but there is a lack of empirical evidence as to how this can be successfully implemented in a way that engages relevant stakeholders.

This thesis therefore sought to address three principal research questions:

1. How do students engage with feedback in a high stakes context?
2. What are the influences and barriers to feedback receptivity, both at an individual and institutional level, in this context?

3. Can we change the assessment culture to one which prioritises and supports learning from feedback? What is stopping us?

The study described in **Chapter 2** sought to address the first main research question by using quantitative methodology to evaluate the extent to which medical students engaged with feedback after a high stakes clinical assessment. To address the problems of limited faculty resources and the need for timely feedback delivery, feedback information was provided in a web-based format. As well as helping with feedback delivery, the web-based format also enabled students' use of the information to be monitored. This website usage data was linked to students' performance in the assessment and their responses to a questionnaire concerning learning-related student characteristics. As an additional feature, the website provided guidance to learners on formulating an action plan following study of the feedback. Our analysis demonstrated that most students engaged with web-based feedback after a summative OSCE, but at a very superficial level. While nearly all students visited the website, just over half viewed the information only once. Many students looked at each separate webpage for a very limited time, with half the students viewing at least 6 webpages per minute. The webpages for each OSCE station which focussed on students' proximity to the pass-fail mark were visited much more frequently than those pages that gave a much more detailed breakdown of how the student performed within the stations, or in their performance across stations. Very few students viewed the 'next steps' pages which gave students guidance in how to use the information. Use of the feedback website was not homogenous, with a wide range in the number of webpages viewed and the number of times the website was visited. The students who had only just passed the OSCE were the least engaged; none visited the website more than twice, even though they would potentially have much to gain from feedback. Although the students who passed all stations typically looked at more webpages, their profile of usage was not particularly reassuring; they made more use of aspects of the website which compared their performance with other students, rather than looking at the pages which demonstrated why they have performed strongly. The conclusion from this study was that it is not enough simply to provide feedback and expect students to make full use of the opportunities provided. While this quantitative study provided useful information on how students engage with feedback in a summative context, it failed to provide an explanation of the reasons why learners may, or may not, engage with feedback offered in these circumstances.

The study in **Chapter 3** therefore employed in-depth interviews to explore the individual barriers to feedback receptivity. One of the main findings was the perceived power of the summative assessment culture. The fear of failure and the subsequent "punishment" of needing resits loomed large in students' minds. Learners felt that more attention was focussed on those students who failed, while students who passed felt ignored. This meant that exams were seen as a hurdle, not a stepping stone to future learning. Additionally, students experienced a rollercoaster of strong emotions: fear of taking the exam, relief when it's over, fear of getting the results, intense relief when passed. Interactions with

Summary

others reinforced the need to avoid failure rather than use the feedback information as an aid towards improved clinical performance. Teachers talked about it being a hurdle unrelated to real life as a doctor. Discussions between students focussed on whether they needed to resit or not. Students tended to share the feedback information more with parents, partly to seek approval. Parents reinforced the need to just pass. Students had been used to succeeding in exams at school, but they now had to adjust to not coming top of the class. This led to a sense of futility, which led in turn to an attitude of not trying too hard, and instead focussing on just passing the assessment. There was a disconnection between assessment and future clinical learning. Feedback was seen as relevant for a specific exam or a specific question; it was felt that it might help performance in future exams but was rarely seen as relevant for clinical workplace. Clinical teachers rarely discussed feedback from assessment to help students apply learning in workplace. Overall, students did not appear motivated to make use of feedback.

Although the study in chapter 3 set out to explore the individual barriers to feedback receptivity, it demonstrated that factors within a medical school's assessment culture were influential. The study in **Chapter 4** therefore sought to understand the factors within medical school assessment systems in more detail. Focus groups were held in three different medical schools with very different assessment cultures. Assessment systems either increased or decreased personal agency. Receptivity to feedback was increased if agency was increased. Within assessment, agency was increased by providing elements of choice, enabling students to demonstrate their knowledge to themselves or their tutors. Within feedback, agency was enhanced by enabling students to act independently in accessing feedback. Institutions reduced agency by controlling how feedback was used, which led students to play an 'assessment game'. Students recognised the need for limits to agency. Standardisation of assessments was seen as fair, as long as the assessment had relevance to their work as future doctors. Assessments were often perceived as irrelevant to future clinical work as a doctor, which had two adverse effects. Feedback received in the workplace before the assessment was ignored, as it would not help students to pass the assessment. Secondly, feedback received following the assessment was not seen as relevant for clinical learning. Grades provided reassurance and clarity to students that they were on track. But they did not act as an incentive for students to aim higher – rather the reverse, as high grades seemed unrealistic to obtain. Not having grades caused uncertainty, but was seen as a more authentic preparation for real life as a doctor and allowed a more nuanced view of students' strengths and weaknesses. The uncertainty helped to drive students to aspire towards achieving optimal competence. Mentoring helped students to interpret feedback and could support the fostering of agency if the scaffolding was reduced as time went on through the course. A long-term relationship with a mentor enabled trust to develop. Mentors could see all the feedback relating to the student, which enabled new feedback to be put into context. They could challenge students' inaccurate self-assessments.

These studies demonstrated how assessment cultures have remained rooted in a behaviourist approach to medical education: students are rewarded for passing and punished for failing assessments. Furthermore, the studies supported a move towards a more constructivist approach to assessment, empowering the students to take a more active role. Other constructivist elements were a clear call for more authenticity in assessments and the benefits of mentoring, with gradual descaffolding as the course progresses.

The previous studies had highlighted the need to change the assessment culture in order to enhance feedback receptivity. The study described in **Chapter 5** used participatory design methods to explore how a mixed group of stakeholders might attempt to change the assessment culture to one in which students made more use of post-assessment feedback. There was recognition that current assessments could be inauthentic, so that performance in assessments did not necessarily reflect clinical workplace learning. As a result, inauthentic assessments could reward last-minute learning away from the workplace. There was often a lack of time or opportunity for students to demonstrate their knowledge and skills. There was also recognition that feedback was often ignored if the assessments were inauthentic. However, implementation of more authentic assessment was seen as very challenging, as increasing an assessment's authenticity would lead to a reduction in its reliability. Participants had experienced efforts to improve the authenticity of individual assessments and found these attempts to be often clumsy and unhelpful. Current inauthentic assessments were seen as having some advantages in that they prepared learners for similar assessments early in their postgraduate career.

The provision of feedback without grades was considered but was perceived to be very difficult to implement for a number of reasons. Firstly, it was felt that learners would attempt to convert narrative feedback into a numerical mark or grade even if none was offered. This in turn could be misleading, as there was felt to be a mismatch between current narrative feedback and the numerical mark; students with a mark close to the pass-fail boundary often received quite encouraging feedback, while excellent students often received quite critical feedback with the aim of stretching the learner. Examiners felt that the numerical mark was more important than the narrative feedback. Grades were felt to provide students with clarity and reassurance. Despite awareness of the literature supporting the provision of feedback without grades, participants did not believe this evidence as it did not resonate with their own personal experience.

Mentoring was seen as a beneficial way to encourage students to engage meaningfully with feedback but was also regarded as presenting enormous implementation challenges. While it was recognised that mentoring needed to occur soon after the provision of assessment-related feedback, enough time needed to be allowed to ensure the process was meaningful. Similarly, selection of appropriate mentors provided other practical challenges, with a tension between the need for a long-term relationship with the tutor versus the credibility of feedback coming from a subject specialist.

Summary

There was evidence that personal and collective beliefs were influential in ensuring that participants were reluctant to change. There was a strong shared belief in the primacy of the summative assessment paradigm, which meant that the pass-fail cut score, and the need to exclude unsafe doctors, was seen as the most important purpose of assessments. Allied to this was a need for objectivity and rigour in order for the process to be fair. Participants' prior experiences as learners reinforced their belief in the summative assessment. However, when participants had experienced a difference assessment culture, there was openness to change.

Within the group setting, senior members in particular often expressed their beliefs in a tacit or implicit manner. Beliefs were expressed more explicitly in the context of individual interviews, with the result that resistance to change was not fully apparent from the group setting. Beliefs did not appear easily influenced by the comments of others during the group discussion. Despite dissatisfaction with many aspects of the current assessment culture, and recognition of a need for more meaningful engagement with feedback, there appeared to a shared common assumption that "summative assessment is the way things are done round here" with a result that summative assessment was seen as more important than feedback. This study demonstrated the enormity of the challenge for assessment programme designers who wish to bring about a change in assessment culture.

Finally **Chapter 6** considered the answers to the research questions in the context of the wider research literature. The challenging conclusion was that, although the practical difficulties of delivering feedback in summative assessment context can be overcome, learners will often not be receptive to the feedback provided, as the summative assessment culture is dominant and stifles the learning potential available from assessments. However, when potential change of assessment culture was considered, faculty and students were very reluctant as they maintained strong beliefs in the benefits of summative assessment, often because of their own prior experiences. Despite this, the weight of educational evidence suggests that maintaining the status quo is not an acceptable solution. The thesis concludes by calling for a paradigmatic shift in assessment and feedback culture. When designing a programme of assessment, medicine faces particular challenges because of the need for regulators to reassure the public that their doctors are competent. In order to assist clinician-educators to make the leap of faith towards an assessment culture which fosters learning, analogies are drawn from the field of healthcare. Beliefs in assessments have similarities with patients' beliefs about illness. The thesis argues that these beliefs should be explored in a way that fosters mutual respect, to try to reach compromises which are reasonably acceptable to all stakeholders within an assessment programme. While accepting the significant challenges of implementing long-term mentoring, the thesis argues that there is great potential that clinicians and students would subsequently find the assessment system more rewarding in the end, in much the same way that clinicians and patients find continuity of care extremely beneficial. After considering the strengths and limitations of the thesis, suggestions are made for further research. Although the need for

coherent, well-designed assessment programmes is increasingly accepted, it is essential to research more closely the actual implementation of change in the assessment culture. Suggestions for how this may be achieved are considered.

Samenvatting

Feedback in de context van ‘*high stakes*’-toetsing: kan summatief formatief zijn?

Feedback kan een zeer krachtig middel zijn; een van de meest krachtige manieren om de prestaties te bevorderen. Helaas wordt feedback vaak niet optimaal benut. Dit komt bijvoorbeeld doordat de feedback niet op de juiste wijze wordt gegeven of ontvangen of doordat ontvangers niet weten hoe zij de feedback moeten vertalen naar concrete leeractiviteiten die tot positieve veranderingen leiden. Er is tot op heden weinig aandacht besteed aan de rol van feedback in de context van summatieve toetsing. Het doel van dit proefschrift was dan ook om stil te staan bij de moeilijkheden die bij het ontvangen van feedback in deze context komen kijken, en meer specifiek om de gedeelde opvatting dat toetsen een formatief dan wel een summatief doel hebben en dat het lastig is om beide doelen in één enkele toets te verenigen ter discussie te stellen.

Er zijn verschillende redenen om aan te nemen dat het verstandig zou zijn om de formatieve en summatieve doelen van toetsing met elkaar te verenigen. ‘*High stakes*’-toetsen genereren doorgaans grote hoeveelheden gegevens over de sterke en zwakke punten van een student, maar deze informatie wordt teruggebracht tot een binair voldoende/onvoldoende-besluit over bevordering naar de volgende fase. Hoewel deze aanpak begrijpelijk is, kan dit misleidend zijn, aangezien zo veel waardevolle informatie over de student in het proces verloren gaat; bovendien bereidt een dergelijke aanpak de studenten niet voor op zelfstandig zelfsturend leren. Gedetailleerde informatie uit toetsen kan studenten helpen inzicht te verkrijgen in hun eigen sterktes en aandachtspunten. Praktijktoetsen zijn zeer duur in het medisch onderwijs, dus vanuit financieel oogpunt is het ook zinvol om zoveel mogelijk voordeel uit deze toetsen te behalen.

Aan de andere kant brengt het geven van feedback in deze omstandigheden praktische knelpunten met zich mee, zoals een behoefte aan een verhoogde inzet van staf, hoewel dit deels met behulp van informatietechnologie mogelijk ondervangen kan worden. De ontvankelijkheid van studenten voor feedback vormt echter een groter aandachtspunt; het is niet zeker of zij in deze context op een zinvolle manier met de feedback aan de slag zullen gaan. Wellicht zijn er individuele of cultuurgebonden factoren die de ontvankelijkheid voor feedback belemmeren, maar hier is geen onderzoek naar gedaan in de context van summatieve toetsing. Het is bekend dat summatieve toetsing andere aspecten van het leerproces nadelig beïnvloedt, dus het is aannemelijk dat deze toetswijze ook ongewenste gevolgen kan hebben voor de ontvankelijkheid voor feedback. Hoewel er voorstellen zijn gedaan om de summatieve toetscultuur om te vormen tot een cultuur waarbij de nadruk ligt op toetsen om te leren, ontbreekt er wetenschappelijk bewijs dat aantoont hoe dit zodanig kan worden ingevoerd dat belanghebbenden er ook daadwerkelijk bij betrokken raken.

Dit proefschrift beoogde daarom drie centrale onderzoeksvragen te beantwoorden:

1. Hoe gaan studenten aan de slag met feedback in een ‘*high stakes*’-situatie?
2. Welke individuele en instellingsgebonden factoren beïnvloeden en belemmeren de ontvankelijkheid voor feedback in een dergelijke situatie?

3. Kunnen we de toetscultuur omvormen tot een cultuur waarin leren van feedback bevorderd wordt en voorop staat? Wat houdt ons tegen?

De in **Hoofdstuk 2** beschreven studie beoogde de eerste centrale onderzoeksvraag te beantwoorden door met behulp van kwantitatieve methodes te beoordelen in welke mate geneeskundestudenten aan de slag gingen met feedback na een ‘*high stakes*’-stationstoets. Om de inzet van staf zoveel mogelijk te beperken en om snelle levering van de feedback te faciliteren, werd de feedbackinformatie elektronisch via het web beschikbaar gesteld. De digitale format maakte het niet alleen eenvoudiger om feedback te geven, maar registreerde ook hoe studenten de informatie gebruikten. De gegevens betreffende het gebruik van de website waren gekoppeld aan de resultaten die studenten behaald hadden op de toets en aan hun antwoorden op een vragenlijst over studenteigenschappen met betrekking tot leren. Een aanvullende functionaliteit was dat de website studenten hulp bood bij het opstellen van een actieplan na het bestuderen van de feedback. Uit onze analyse bleek dat de meeste studenten iets deden met de digitale feedback die zij na afloop van een summatieve stationstoets ontvingen, zij het op een erg oppervlakkig niveau. Hoewel nagenoeg alle studenten de website hadden bezocht, bekeek net iets meer dan de helft de informatie slechts één keer. Veel studenten bekeken elke afzonderlijke internetpagina slechts heel kort, en de helft van hen bekeken minstens 6 internetpagina’s per minuut. De pagina’s die per station aantoonde hoe ver de student van een voldoende verwijderd was werden veel vaker bezocht dan de pagina’s die een veel meer gedetailleerde specificatie gaven van hoe de student gepresteerd had binnen de stations en over de diverse stations. Erg weinig studenten bekeken de pagina’s met vervolgstappen die studenten op weg hielpen om iets met de informatie te doen. Het gebruik van de feedbackwebsite verschilde erg van student tot student: er was een grote spreiding in het aantal internetpagina’s dat zij bekeken en het aantal keer dat zij de website bezochten. De studenten die de stationstoets met een nauwe marge hadden behaald waren het minst betrokken; geen van hen bezocht de website meer dan twee keer, ondanks dat zij mogelijk veel baat zouden hebben van de feedback. Hoewel de studenten die voor alle stations geslaagd waren doorgaans meer internetpagina’s bekeken, was hun gebruiksprofiel niet bijzonder geruststellend; in plaats van te kijken naar de pagina’s die lieten zien waarom zij zo goed gepresteerd hadden, maakten zij meer gebruik van aspecten van de website die hun prestaties met die van andere studenten vergeleken. De conclusie uit deze studie was dat het niet volstaat om studenten eenvoudigweg van feedback te voorzien en te verwachten dat zij de gegeven kansen volledig benutten. Hoewel deze kwantitatieve studie zinvolle informatie heeft verschaft over hoe studenten in een summatieve context omgaan met feedback, gaf het geen verklaring van de redenen waarom studenten wel of niet van feedback gebruik maken in deze omstandigheden.

De studie in **Hoofdstuk 3** maakte daarom gebruik van diepte-interviews om te onderzoeken welke individuele factoren de ontvankelijkheid voor feedback belemmeren. Een van de belangrijkste bevindingen was de waargenomen invloed van de summatieve toetscultuur.

Angst om te falen en voor de daaruit volgende “straf” te moeten herkansen hing als een donkere wolk boven het hoofd van de studenten. Zij hadden het gevoel dat de aandacht voornamelijk uitging naar de studenten die de toets niet hadden gehaald, terwijl de studenten die waren geslaagd zich gepasseerd voelden. Met andere woorden, de examens werden eerder gezien als een horde dan als opstapje naar toekomstig leren. Bovendien hadden studenten het gevoel zich in een achtbaan van sterke emoties te bevinden: angst voor het afleggen van het examen, opluchting wanneer het voorbij was, angst voor het krijgen van de uitslag, grote opluchting wanneer ze waren geslaagd. Interacties met anderen versterkten de behoefte om falen te voorkomen in plaats van dat de daaruit vloeiende feedbackinformatie werd aangewend als hulpmiddel voor het verbeteren van de klinische prestaties. Docenten spraken van een horde die niets te maken had met het echte leven van een arts. Gesprekken tussen studenten gingen er voornamelijk over of ze wel of niet moesten herkansen. Studenten hadden de neiging om de feedbackinformatie meer met de ouders te delen, deels omdat ze goedkeuring zochten. Ouders versterkten de behoefte om gewoon te slagen. Studenten waren gewend met glans te slagen voor toetsen op school, maar nu moesten ze zich aanpassen aan het feit dat ze niet meer de beste van de klas waren. Als gevolg hiervan voelden zij zich kansloos, wat er weer toe leidde dat studenten zich niet al te veel inspanden en zich in plaats daarvan richtten op het simpelweg behalen van de toets. Er was geen verband tussen toetsing en toekomstig klinisch leren. Feedback werd gezien als relevant voor een specifieke toets of specifieke vraag; men zag het nut ervan in voor het eventueel verbeteren van prestaties op toekomstige toetsen, maar zelden voor de klinische werkplek. Klinische docenten bespraken de toetsfeedback zelden met studenten om hen te helpen het leren op de werkplek toe te passen. Over het algemeen leken de studenten niet gemotiveerd om van de feedback gebruik te maken.

Hoewel het doel van de studie in hoofdstuk 3 was om te onderzoeken welke individuele factoren de ontvankelijkheid voor feedback belemmeren, werd ook aangetoond dat factoren met betrekking tot de toetscultuur van een medische faculteit een belangrijke rol spelen. **Hoofdstuk 4** poogde daarom meer licht te werpen op de factoren binnen toetssystemen van medische faculteiten. Er werden focusgroepen gehouden in drie verschillende medische faculteiten met zeer uiteenlopende toetsculturen. Toetssystemen hadden ofwel een versterkend dan wel een beperkend effect op de persoonlijke invloed. De ontvankelijkheid voor feedback nam toe naarmate er meer verantwoordelijkheid bij de student werd gelegd. Binnen toetsing werd de persoonlijke invloed verhoogd door sommige keuzes aan de student te laten, zodat zij hun kennis aan henzelf of aan hun tutoren konden laten zien. Op het niveau van feedback werd de persoonlijke invloed versterkt door studenten in staat te stellen de feedback op zelfstandige wijze te verkrijgen. Instellingen beperkten de invloed van studenten door te bepalen hoe de feedback werd gebruikt, waardoor de studenten een “toetsspel” gingen spelen. De studenten zagen in dat er grenzen gesteld moesten worden aan hun invloed. Het standaardiseren van toetsen werd goedgekeurd, zolang de toets relevant was voor hun werk als toekomstig arts. Toetsen werden vaak gezien als irrelevant voor het toekomstige klinisch werk als arts, wat twee nadelen met zich meebracht. In de

eerste plaats deden studenten niets met de feedback die zij vóór de toets op de werkplek ontvingen, omdat deze hen niet hielp om de toets te halen. Ten tweede beschouwden zij de feedback die zij na afloop van de toets ontvingen niet als relevant voor klinisch leren. Cijfers stelden studenten gerust en boden inzicht in de vraag of ze wel of niet bij waren. Ze zorgden er echter niet voor dat studenten hun doelen naar boven bijstelden; het tegenovergestelde was eerder het geval, aangezien het behalen van hoge cijfers onrealistisch leek. Geen cijfers zorgde voor onzekerheid, maar werd gezien als een meer authentieke voorbereiding op het echte leven als arts en bood studenten een meer genuanceerd inzicht in hun sterktes en aandachtspunten. Door de onzekerheid werden studenten aangespoord te streven naar optimale competentie. Begeleiding van een mentor hielp studenten de feedback te interpreteren en kon de persoonlijke invloed versterken, mits de ondersteuning in de loop van de opleiding geleidelijk minder werd. Een langdurige relatie met een mentor bood ruimte voor het opbouwen van vertrouwen. Het stelde mentoren in staat alle op de student van toepassing zijnde feedback in te zien, waardoor nieuwe feedback in context kon worden geplaatst. Zij konden de studenten confronteren met hun eventuele foute zelfbeoordelingen.

Deze studies toonden aan hoe toetsculturen nog steeds verankerd zijn in een behavioristische benadering van medisch onderwijs: studenten worden beloond wanneer zij de toetsen halen en gestraft wanneer zij deze niet halen. Verder schraagden de studies een overgang naar een meer constructivistische benadering van toetsing, waarbij studenten in staat worden gesteld een actievere rol te vervullen. Andere constructivistische aspecten waren een duidelijke oproep tot meer authenticiteit in toetsen en de voordelen van begeleiding door een mentor, met geleidelijke afbouw van ondersteuning in de loop van de opleiding.

De voorgaande studies hebben benadrukt dat de toetscultuur eerst moet veranderen wil men de ontvankelijkheid voor feedback vergroten. De in **Hoofdstuk 5** beschreven studie onderzocht met behulp van participatief-ontwerpmethodes hoe een gemengde groep belanghebbenden de toetscultuur zou proberen om te vormen tot een cultuur waarin studenten meer gebruik maken van de feedback die zij na de toets ontvangen. Er werd toegegeven dat de huidige toetsen mogelijk niet authentiek waren, zodat de prestaties op toetsen niet per definitie een weerspiegeling gaven van het leren op de klinische werkplek. Hierdoor kon het zijn dat niet-authentieke toetsen leren op het laatste moment buiten de werkplek beloonden. Studenten hadden vaak te weinig tijd of kansen om hun kennis en vaardigheden te tonen. Ook werd erkend dat de feedback vaak in de wind werd geslagen wanneer toetsen niet authentiek waren. De invoering van meer authentieke toetsen werd echter als iets zeer lastigs gezien, omdat een toename in de authenticiteit gepaard zou gaan met een afname van de betrouwbaarheid van de toets. Deelnemers hadden pogingen meegemaakt om de authenticiteit van afzonderlijke toetsen te verbeteren, maar vonden deze vaak onhandig en nutteloos. Een van de voordelen die zij in de huidige niet-authentieke

toetsen zagen was dat deze de studenten voorbereidden op vergelijkbare toetsen die zij in het begin van hun vervolgopleiding zouden krijgen.

De optie om feedback te geven zonder cijfers werd overwogen, maar dit werd om allerlei redenen gezien als zeer lastig om in te voeren. In de eerste plaats was men de mening toegedaan dat zelfs wanneer er geen punt gegeven werd studenten zouden proberen om de narratieve feedback te vertalen naar een numerieke beoordeling of cijfer. Dit kon misleidend zijn, aangezien men vond dat de huidige narratieve feedback en numerieke beoordeling niet met elkaar overeenstemden. Studenten die met hun cijfer dicht bij de voldoende/onvoldoende-grens zaten ontvingen namelijk vaak redelijk aanmoedigende feedback, terwijl de feedback die uitmuntende studenten ontvingen vaak erg kritisch was met de bedoeling de student nog meer uit te dagen. Examinatoren vonden de numerieke beoordeling belangrijker dan de narratieve feedback. Men vond dat cijfers de studenten duidelijkheid verschafte en geruststelden. Ondanks dat deelnemers ervan op de hoogte waren dat het geven van feedback zonder cijfers door de literatuur werd ondersteund, geloofden zij niet in dit bewijs omdat het niet strookte met hun eigen persoonlijke ervaring.

Begeleiding van een mentor werd gezien als een nuttige manier om studenten aan te moedigen op een zinvolle manier met de feedback om te gaan, maar tegelijkertijd werden grote problemen voorzien met de invoering ervan. Hoewel men inzag dat mentorbegeleiding zo snel mogelijk na het geven van toets-gerelateerde feedback diende plaats te vinden, moest er ook voldoende tijd worden geboden om de zin van het proces te kunnen garanderen. Ook kwamen er andere praktische problemen bij de selectie van geschikte mentoren om de hoek kijken, bijvoorbeeld omdat de behoefte aan een langdurige relatie met de tutor botste met de geloofwaardigheid van feedback afkomstig van een vakspecialist.

Er was bewijs dat persoonlijke en gemeenschappelijke overtuigingen in grote mate garandeerden dat deelnemers van verandering afkerig waren. Er heerste een intense, gedeelde overtuiging dat het summatieve toetsmodel superieur was, wat inhield dat de voldoende/onvoldoende-grensscore, en de behoefte om onveilige arts en artsen uit te sluiten, als belangrijkste doel van de toetsen werd gezien. Hieraan verbonden was een behoefte aan objectiviteit en zorgvuldigheid om een eerlijk proces te kunnen garanderen. Het geloof in de summatieve toets werd versterkt door de ervaringen die de deelnemers zelf hadden gehad als student. Wanneer de deelnemers echter een andere toetscultuur hadden ervaren, stond men open voor verandering.

In groepsverband spraken de oudere leden hun overtuigingen vaak op een stilzwijgende of impliciete manier uit. Tijdens de individuele interviews werden de overtuigingen explicieter uitgesproken, waardoor uit het groepsverband niet helemaal duidelijk werd dat zij van verandering afkerig waren. Overtuigingen bleken niet gemakkelijk beïnvloed te worden door de opmerkingen van anderen tijdens de groepsdiscussie. Ondanks dat er

ontevredenheid bestond over vele aspecten van de huidige toetscultuur en men inzag dat studenten meer op een zinvolle manier aan de slag moesten gaan met de feedback, bleek er een gedeelde opvatting te heersen dat “summatieve toetsing is zoals we het hier aanpakken”, waardoor summatieve toetsing als belangrijker werd beschouwd dan feedback. Deze studie toonde aan hoe immens groot de uitdaging is waarvoor programmaontwikkelaars die een verandering in toetscultuur willen doorvoeren, worden gesteld.

Ten slotte werden in **Hoofdstuk 6** de antwoorden op de onderzoeksvragen beschouwd in de context van de bredere onderzoeksliteratuur. De prikkelende conclusie was dat, ook al is het mogelijk de praktische moeilijkheden die het geven van feedback in een summatieve toetscontext met zich meebrengt te boven te komen, studenten vaak niet open zullen staan voor de feedback, omdat de summatieve toetscultuur overheerst en deze de leerkansen die toetsen te bieden hebben in de kiem smoort. Echter, wanneer een mogelijke verandering in toetscultuur werd overwogen, waren stafleden en studenten erg afkerig omdat zij hevig geloofden in de voordelen van summatieve toetsing, meestal door hun eigen eerdere ervaringen. Desalniettemin weegt het onderwijskundige bewijs te zwaar om te opereren dat het handhaven van de status quo een aanvaardbare oplossing is. Het proefschrift sluit af met een oproep tot een paradigmaverschuiving in toets- en feedbackcultuur. Bij het ontwerpen van een programma van toetsing doen zich in het geval van geneeskunde specifieke moeilijkheden voor omdat er regulerende maatregelen nodig zijn om de mensen te verzekeren dat hun artsen competent zijn. Om arts-opleiders te helpen bij de vertrouwenssprong naar een toetscultuur die het leren bevordert, worden vergelijkingen getrokken met voorbeelden uit de gezondheidszorg. Toetsovertuigingen vertonen gelijkenissen met opvattingen van patiënten over ziekte. Dit proefschrift stelt dat deze overtuigingen onderzocht zouden moeten worden op een manier die wederzijds respect stimuleert, zodat men kan proberen compromissen te sluiten die voor alle belanghebbenden in een toetsprogramma aanvaardbaar zijn. Hoewel het proefschrift onderkent dat de invoering van langdurige begeleiding door een mentor behoorlijk wat moeilijkheden met zich meebrengt, stelt het dat de kans groot is dat klinici en studenten het toetsstelsel uiteindelijk meer zullen gaan waarderen, vergelijkbaar met de manier waarop klinici en patiënten de continuïteit van de zorg uiterst zinvol achten. Na een beschouwing van de sterke en zwakke punten van het proefschrift worden suggesties gedaan voor toekomstig onderzoek. Hoewel het belang van samenhangende, goed ontwikkelde toetsprogramma's steeds meer wordt ingezien, is het essentieel dat de daadwerkelijke invoering van verandering in de toetscultuur nader wordt onderzocht. Er wordt een uiteenzetting gegeven van mogelijke manieren om dit te bereiken.

