

Nut en waarde van multiple choice toetsen

Bert ten Boske (Faculteit Techniek/HvA), Veronica Bruijns (Onderwijs en Onderzoek/HvA)

1 Inleiding

Sinds september 2015 beschikken HvA en UvA over twee grote digitale toetszalen met een capaciteit van 600 studenten. Opleidingen worden gestimuleerd van deze zalen gebruik te maken voor de afname van digitale tentamens. Dit zijn overwegend multiple choice toetsen. Dat maakt binnen de HvA de discussie (weer) actueel wat het nut en de waarde van multiple choice toetsen is in onze bacheloropleidingen. Is het wel de juiste toetsvorm voor de beoordeling van de ontwikkeling van de student? Volgens Nieweg (2014) is een multiple choice toets geen neutraal instrument; het weerspiegelt de opvatting over leren en lesgeven. Anderen zien het als een tijdloze toetsvorm, die goed inzetbaar is bij grootschalig onderwijs.

Met dit artikel beogen we om docenten/ontwikkelaars van multiple choice toetsen te informeren over actuele kennis over nut en waarde van gesloten vragen. We starten met een samenvattend advies in paragraaf 2. In paragraaf 3 geven we op basis van de onderzoekliteratuur de voor- en nadelen van multiple choice toetsen. Paragraaf 4 bevat tips voor de constructie en de inzet van multiple choice vragen. De onderbouwing van de onderzoeksliteratuur is beschreven in paragraaf 5. De bijlage biedt een schematisch overzicht van de opbrengst uit de onderzoeksliteratuur.

Leeswijzer: paragraaf 2 t/m 4 zijn bedoeld voor wie snel terzake wil komen. Paragraaf 5 en de bijlage zijn toegevoegd voor wie geïnteresseerd is in de onderzoeksinformatie waarop de voor- en nadelen en tips gebaseerd zijn.

2 Advies voor gebruik multiple choice vragen

Het overall advies voor het gebruik van multiple choice toetsen op basis van de literatuur: indien zorgvuldig samengesteld vanuit een toetsmatrijs, en met inachtneming van de niveaus van de Bloom taxonomie en de regels voor constructie, is de toetsvorm goed inzetbaar voor grote groepen studenten. Vermijdt vragen die op reproductie zijn gericht. Construeer vragen die begrip en toepassing van relevante kennis toetsen, wetende dat dit meer tijd en expertise vraagt. Zorg, als opleiding gezamenlijk, voor een kwalitatief goede opbouw van het toetsprogramma, met een evenwichtige mix van toetsvormen, gebaseerd op de te verwerven competenties. Maak optimaal gebruik van de formatieve mogelijkheden van multiple choice toetsen, als effectieve ondersteuning van het leerproces van student én docent. Vergroot de betrokkenheid van de studenten door gebruik te maken van vernieuwende praktijken in het gebruik van multiple choice toetsen.

Noot bij dit advies: we hebben niet de illusie dat we, met de actuele kennis uit onderzoeksliteratuur, de discussie over nut en waarde van multiple choice toetsen voor eens en altijd kunnen beslechten. Daarvoor zijn de resultaten uit de onderzoeken te wisselend. Het advies, op basis van de onderzoekliteratuur, is die van de auteurs. We nodigen lezers uit, om op basis van de bevindingen uit paragraaf 5, hun eigen conclusies te trekken.

3 Voor en nadelen multiple choice toetsen

Op basis van de onderzoeksliteratuur komen we tot de volgende voor-en nadelen van kwalitatief goed samengestelde multiple choice toetsen:

3.1 Voordelen

1. Multiple choice toetsen zijn kosteneffectief ten aanzien van het beoordelingsproces; correctie van gesloten vragen kost relatief weinig tijd.
2. Bij multiple choice toetsen is er geen sprake van bias of beoordelingsfouten bij de eindbeoordeling. De beoordeling is objectief; bij essays is deze afhankelijk van de interpretatie van de docent/examinator.
3. De validiteit en betrouwbaarheid zijn bij multiple choice toetsen gewaarborgd en (achteraf) goed te bepalen.
4. Bij multiple choice toetsen kan direct feedback worden gegeven, zowel bij summatieve (beoordeling) als bij formatieve toetsing (ondersteuning leerproces)
5. Multiple choice toetsen maken het mogelijk om een breed scala aan onderwerpen in één keer te toetsen op verschillende cognitieve niveaus.
6. Multiple choice toetsen kunnen hogere orde vaardigheden toetsen, hoewel de constructie hiervan veel expertise en aanzienlijke tijdsinvestering vereist.
7. Bij een goede opbouw van de toets verkrijgt de docent informatie over waar het begrip c.q. inzicht van de student stopt en waar aandacht aan moet worden besteed.
In het verlengde hiervan: Er kan eenvoudig een uitgebreide kwaliteitsanalyse worden uitgevoerd; de toepassing Learning Analytics (LA)
Voor de opbouw dient de taxonomie van Bloom als uitgangspunt:
 1. Vragen die kennis toetsen; Bloom niveau 1
 2. Vragen die begrip toetsen; Bloom niveau 2
 3. Vragen die toepassing toetsen; Bloom niveau 3
 4. Vragen die analytisch vermogen toetsen; Bloom niveau 4
8. Multiple choice toetsen lenen zich goed voor samenwerking tussen opleidingen en instellingen bij de opbouw van een itembank. Kennis, inzichten en deskundigheid op het gebied van ontwerp kunnen worden gedeeld.
9. Met het gezamenlijk opzetten van de itembank kan literatuur worden losgelaten; toetsontwerp vindt plaats op basis van onderwerpen, leerdoelen, toetstermen. Er is consensus over de gehanteerde terminologie.
10. De gezamenlijk opgestelde itembank kan snel en eenvoudig up-to-date worden gehouden.
11. Bij kwalitatief goede vragen ondervinden taalzwakke kandidaten geen nadeel; taalvaardigheid wordt niet (onbedoeld) meegetoetst.

3.2 Nadelen

1. Het construeren van multiple choice toetsen voor het beoordelen van hogere orde vaardigheden vereist veel expertise en tijd. Dit doel is eenvoudiger te bereiken met de constructie van open vragen en opdrachten.
2. De gesloten vraagvorm met gestandaardiseerde antwoorden biedt geen ruimte voor eigen interpretatie/invulling/creativiteit en beoordeling van het denkproces van de student. Het nodigt uit tot gokken.
3. Bloom's hoogste niveaus - creëren en evaluatie - kunnen niet gemeten worden met de multiple choice toets, evenals creativiteit en originaliteit.

4. Studenten associëren multiple choice vragen met reproductie en herkenning en zijn geneigd een oppervlakkige leerstijl toe te passen bij het voorbereiden op multiple choice toetsen.
5. Het goed beantwoorden van multiple choice vragen kan ten onrechte de indruk geven dat studenten de materie begrepen hebben.
6. Voor meerdere toetsafnames is een grote en dus arbeidsintensieve itembank vereist.
7. Meerdere afnames vereist een grote itembank. Het opstellen van een toets databank is tijdrovend en complex.
8. Studenten worden geconfronteerd met meerdere foute antwoorden, die als juist worden gepresenteerd. Uit onderzoek blijkt dat wanneer studenten later opnieuw bevroegd worden ze foute antwoorden als juist herinneren (Roediger, 2005).

4 *Tips voor docenten/ontwikkelaars*

4.1 *Vooraf: stel een toetsmatrijs samen.*

Een toetsmatrijs is een blauwdruk van een toets waarin wordt aangegeven wat precies getoetst moet worden, met welk doel, op welk niveau en in welke vorm. De matrijs is een hulpmiddel voor het bepalen van een juiste verdeling van de vragen over de toetstermen (leerdoelen) en het aangeven van de cognitieve- en gedragsniveaus.

Aan de hand van de toetsmatrijs wordt de “toetsmix” samengesteld; een mix aan vraagsoorten en toetsvormen. Daarmee worden de sterke punten van de vraagsoorten benut en de zwakke punten geneutraliseerd. De toets wordt daarmee representatief voor de te toetsen stof.

4.2 *Bepaal of de toetstermen/leerdoelen met gesloten vragen beoordeeld kunnen worden*

Bestudeer de toetstermen die volgens de toetsmatrijs moeten worden getoetst. Selecteer vervolgens de vraagvorm die past bij de toetsterm. Naast meerkeuze vragen zijn digitaal verschillende vraagvormen mogelijk, zoals:

- meer- uit meer-vraag;
- numerieke vraag;
- sorteervraag;
- combineervraag;
- matrix- en de matrixkeuzevraag (pull-downlist);
- sleepvraag;
- aanwijsvraag.

Voor nadere toelichting over deze vraagvormen, zie: www.score.hva.nl/docent/digitaal/toetsen.

4.3 *Bepaal op welk niveau van de taxonomie van Bloom de multiple choice vragen betrekking moeten hebben*

Uit de literatuur blijkt dat vooral de eerste drie niveaus van Bloom goed kunnen worden afgedekt met multiple choice vragen. Het vierde niveau - analyseren - kan in gesloten vraagvorm gevat worden, maar vereist grote deskundigheid bij het ontwerpen ervan.

4.4 *Het ontwerpen van de vraag*

- Hanteer een stappenplan voor het ontwikkelen van gesloten vragen;
- Formuleer eerst de stam van de vraag als een vraag die ook zonder antwoordmogelijkheid is te beantwoorden.
- Formuleer eerst het juiste antwoord op de vraag en daarna (indien van toepassing) de afleiders. Maak bij het formuleren van de onjuiste afleiders gebruik van misconcepties ten aanzien van het onderwerp.
- Formuleer feedback op het juiste antwoord en op de onjuiste afleiders (zie verder 3.5).
- Check 1: Leg de vragen na ontwikkeling minimaal een dag weg. Beantwoordt de vragen daarna zonder de alternatieven te bekijken.
- Pas de vraag aan als het antwoord niet gegeven kan worden.
- Check 2: laat de vragen controleren door een collega vraagontwerper.

De schrijfwijzers van ECABO (2012 SURFprojecten “ToetsBabe” en “ToetsenLeer”) geven een houvast voor het formuleren van de toetstermen en het ontwerpen van de vragen en de feedback op de antwoorden. Zie: ECABO, *Schrijfwijzer gesloten vragen*:

<https://score.hva.nl/Bronnen/ECABO-SURF%20-%20Schrijfwijzer%20gesloten%20vragen.pdf>

ECABO, *Schrijfwijzer toetstermen*: <https://score.hva.nl/Bronnen/Schrijfwijzer%20toetstermen%20ECABO-SURF.pdf>

4.5 *Formuleren van de feedback*

- Formuleer de feedback positief om de student te stimuleren in het leerproces.
- Zorg dat de feedback op de onjuiste afleiders inzicht geeft in mogelijke misconcepties.
- Neem geen nieuwe informatie op in de feedback. Dit voorkomt dat studenten het idee krijgen dat ze de vraag niet konden beantwoorden zonder de feedback.
- Geef ook feedback als het antwoord juist is. De student weet dan hij/zij de vraag op de juiste gronden goed heeft beantwoord en/of de onderliggende argumentatie klopt.

Zie de bovengenoemde schrijfwijzers.

4.6 *Gebruik bij het formuleren een checklist om valkuilen te vermijden*

Bij het formuleren van kwalitatief goede gesloten vragen spelen veel aspecten een rol. Door gebruik te maken van een checklist kunnen ontwikkelaars systematisch nagaan of aan alle kwaliteitseisen wordt voldaan. Zie voor een actuele checklist voor gesloten vragen:

https://score.hva.nl/Bronnen/Checklist%20gesloten_vragen.pdf.

4.7 *Benut nieuwe inzichten*

Nicol (2007) geeft aan de hand van de zeven principes van effectieve feedback (Nicol & Macfarlane-Dick, 2006) aan welke rol multiple choice vragen kunnen spelen om de leeropbrengst voor studenten te vergroten:

1. *Geef een duidelijk beeld van de prestatie; (leerdoelen, criteria, standaarden)*. Betrek studenten bij het ontwikkelen van multiple choice vragen. Dit zorgt ervoor dat ze actief betrokken worden bij de leerdoelen en criteria. De beste vragen worden opgenomen in een itembank.

2. *Faciliteer de ontwikkeling van kritische zelfbeoordeling en reflectie gedurende het leerproces.* Gebruik multiple choice vragen als een open boek tentamen. Studenten kunnen zichzelf corrigeren tijdens de test. Het is dan wel van belang dat de vragen op de hogere niveaus van Bloom's taxonomie zijn gericht.
3. *Geef kwalitatief goede informatie over de stand van zaken in het leerproces.* Laat onderwerpen waar studenten slecht op scoorden aan de orde komen in vervolgbijeenkomsten (dit veronderstelt dat de multiple choice toets niet de afsluiting is van een onderwerp).
4. *Stimuleer de dialoog docent-student en student-student over het leerproces.* Laat studenten gezamenlijk multiple choice vragen ontwikkelen. Geef studenten verschillende vragen en laat ze discussiëren over het juiste antwoord.
5. *Stimuleer vertrouwen in eigen kennis en kunnen en gevoel voor eigenwaarde van de student.* Geeft studenten herhaalde mogelijkheden om een multiple choice toets te doen in formatieve vorm. Laat een deel van vragen vervolgens terugkomen in de summatieve toets.
6. *Schep gelegenheid het gat tussen huidige- en gewenste kennis en vaardigheden te dichten.* Geef op meerdere momenten in de studie een multiple choice toets, waarbij de vragen op eindniveau zijn gericht. De voortgangstoetsen bij de opleiding Geneeskunde zijn hiervan een voorbeeld.
7. *Geef de docent informatie over de itemonderwerpen waar extra aandacht aan besteed moet worden.* Stel multiple choice vragen gedurende het leerproces zodat de docent inzicht heeft in de lacunes van de studenten.

Een andere, nieuwe toepassing is *confidence-based marking* (Nicol, 2007). Studenten geven naast het antwoord een score voor hoe zeker ze zijn van hun antwoord. Deze score weegt mee in het eindresultaat.

4.8 *Tijdbesteding ontwerp multiple choice vraag met feedback*

Voor de tijdsbesteding voor het ontwerpen van een toetsvraag met feedback is de volgende richtlijn te geven. Vanuit de ervaringen met de SURF Toetsprojecten ToetsBaBe en Toetsenleer is gebleken dat een ervaren toetsontwerper minimaal 1,5 klokuren nodig heeft om één toetsvraag met 4 alternatieven en met feedback op zowel het goede antwoord als de foute alternatieven te ontwerpen.

5 *Onderbouwing vanuit de onderzoekliteratuur*

Het literatuuronderzoek naar wetenschappelijke artikelen over nut en waarde van multiple choice vragen heeft 23 artikelen opgeleverd (zie bijlage 1). In 13 artikelen zijn, aan de hand van een onderzoeksvraag, multiple choice toetsen vergeleken met andere toetsvormen. 6 artikelen behandelen validiteit en betrouwbaarheid van multiple choice vragen. 4 artikelen gaan in op de leeropbrengst en feedback mogelijkheden. Het aandeel experimentele studies, waarbij een groep studenten een interventie ondergaat en een controlegroep aanwezig is, is beperkt (5). De overige artikelen betreffen case studies of inhoudelijke analyses, veelal op basis van toetsresultaten en enquêtes. Het geringe aandeel experimentele studies betekent dat bij de meeste resultaten sprake is van een verband tussen variabelen, maar dat we niet kunnen spreken van causale relaties.

In bijlage 1 is van ieder artikel een aantal karakteristieken opgenomen.

Uit de artikelen komen de volgende vijf thema's naar voren, die in het vervolg van deze paragraaf nader worden toegelicht aan hand van de resultaten. Bij ieder thema geven we kort onze conclusie, op basis van het literatuuronderzoek. Daarbij houden we rekening met het feit dat in de meeste onderzoeken sprake is van een verband en geen causaal effect:

1. Oppervlakkig versus diepgaand leren in relatie tot multiple choice vragen.

Conclusie: met multiple choice vragen kan kennen, begrijpen, toepassen en analyseren beoordeeld worden. De uitgevoerde onderzoeken leveren nog geen eenduidig beeld op over de mate waarin bestaande multiple choice toetsen deze niveaus ook daadwerkelijk beoordelen. In veel onderzoeken is analyse van de vragen naar de niveau-indeling van Bloom geen onderdeel van het onderzoek. Het is belangrijk om de multiple choice vragen op het juiste niveau te construeren en te zorgen dat er voldoende vragen over het te toetsen onderwerp in de toets zijn opgenomen.

2. De perceptie van studenten ten aanzien van multiple choice vragen.

Conclusie: de perceptie van studenten is dat multiple choice vragen makkelijker zijn dan andere vraagvormen. Om studenten een goed beeld te geven over de aard van de vragen zijn oefentoetsen aan te bevelen.

3. Validiteit en betrouwbaarheid van multiple choice vragen.

Conclusie: vergelijking van multiple choice vragen en open vraagvormen levert geen eenduidig beeld op over de validiteit van multiple choice vragen. Eenmaal goed geconstrueerd heeft de beoordeling van multiple choice vragen een hoge betrouwbaarheid.

4. De onbedoelde effecten van multiple choice vragen.

Conclusie: goede constructie is cruciaal om onbedoelde effecten, zoals testing effect en cueing effect, te minimaliseren.

5. De leerfunctie van gesloten vraagvormen.

Conclusie: het leereffect van multiple choice vragen kan, met eenvoudige interventies, ook bij summatieve toetsing worden verhoogd.

5.1 Oppervlakkig versus diepgaand leren

Een veelvoorkomende opvatting is dat multiple choice toetsen vooral gericht zijn op reproductie van kennis en niet goed in staat zijn om begrijpen en toepassen te beoordelen. In diverse onderzoeken zijn resultaten van studenten op multiple choice vragen en vormen van open of halfopen vragen vergeleken. De resultaten zijn afwisselend positief en negatief. **Walstad en Becker** (1994) constateren dat multiple choice vragen en essay vragen¹ over dezelfde studiestof een overeenkomstig begrip testen van micro en macro-economie. Zij adviseren om essay vragen, gegeven de hogere kosten voor het beoordelen, alleen in te zetten bij een duidelijke toegevoegde waarde. Ook **O'Neill** (2001) vond in een experimentele studie geen invloed op gemiddelde prestaties van studenten op een multiple choice toets en *constructed response test*. **Becker en Johnston** (1999) onderzochten dezelfde vraag als Walstad en Becker. Aan de hand van een solidere statistische analyse, komen ze tot een andere conclusie. De score op een multiple choice toets is volgens hen geen goede voorspeller voor score op essay vragen en andersom. Zij adviseren om beide toetsvormen te gebruiken omdat ze verschillende aspecten toetsen.

Door het geven van het goede antwoord is er nog geen garantie dat de student de materie daadwerkelijk heeft begrepen. **Wooten, Cool, Prather en Tanner** (2014) lieten studenten het antwoord op een multiple choice vraag schriftelijk toelichten. Beiden werden beoor-

¹ Essay vragen zijn, in de onderzoeken die in deze paragraaf zijn opgenomen, open vragen waarbij de student het antwoord - in meer of mindere mate - onderbouwt met argumentatie.

deeld. Studenten presteerden beter op de multiple choice vragen in vergelijking met de aanvul-open-einde-vragen. Zij constateerden dat het kiezen van het juiste antwoord op een multiple choice vraag een overschatting kan zijn van begrip van de student. Ook **Dufresne, Leonard en Gerace** (2002) komen tot dezelfde conclusie. Zij hebben begrip getest in een experiment bij een grote groep studenten, waarbij ze drie samenhangende multiple choice vragen stelden over de wet van Newton, in verschillende volgorde. Uit de resultaten blijkt dat 25% van de studenten alle vragen goed beantwoorden. Hieruit maken ze op dat het goede antwoord op één vraag over bepaalde materie in hoge mate kan berusten op gokkans, intuïtie of onvolledige kennis. Bij het hanteren van één vraag over bepaalde materie kan het juiste antwoord ten onrechte de indruk geven dat de student de materie begrijpt. Zij pleiten daarom voor het gebruik van verschillende toetsvormen om een betrouwbaar beeld te krijgen van competenties van de student.

Hoe kijken studenten hier tegenaan? **Visser, Bootsma en Ten Cate** (2010) constateren, op basis van enquêtes onder medische studenten, dat ze 'een combinatie van drie-, vier- of vijfkeuzevragen en kort-antwoord open vragen' het meest geschikt vinden om studenten te testen op kennis. Bij het toetsen van inzicht en toepassing gaat hun voorkeur uit naar 'een combinatie van drie-, vier- en vijfkeuzevragen, kort-antwoord open vragen en essay vragen' (p. 344). De onderzoekers adviseren docenten om voor het beoordelen van kennis, inzicht en toepassing verschillende typen toetsvragen in één toets te gebruiken.

Een meer essentiële vraag is of met multiple choice toetsen überhaupt hogere orde vaardigheden beoordeeld kunnen worden. **Buckles en Siegfried** (2006) geven, op basis van analyses aan, dat de eerste vier niveaus van de taxonomie van Bloom, - kennen, begrijpen, toepassen en analyseren - in multiple choice vragen te vatten zijn. Het formuleren van vragen op een hoger niveau van de Bloom-taxonomie vereist wel de nodige vaardigheid en tijd. **Palmer en Devitt** (2007) concluderen op basis van een analyse van multiple choice vragen en *modified essay questions* (een kruising tussen multiple choice en essay vragen), dat goed geconstrueerde multiple choice vragen kunnen voldoen aan het cognitieve niveau dat *modified essay questions* vragen beogen. Ook bij *modified essay questions* is constructie de achilleshiel; vragen gericht op reproductie zijn eenvoudiger te construeren dan vragen gericht op begrijpen, toepassen of analyseren.

5.2 Perceptie studenten

De wijze waarop de toetsing is ingericht heeft veel invloed op het leergedrag van studenten. (Gibbs, 2010; Crooks, 1988; Elton & Laurillard, 1979). Hoe kijken studenten tegen multiple choice vragen aan? Vinden ze deze makkelijker dan open vragen? Wat voor invloed heeft het op hun leerstrategie? **Scouller** (1998) heeft de perceptie van studenten onderzocht ten aanzien van multiple choice vragen en essay vragen. Zij concludeert dat studenten oppervlakkig leren toepassen bij multiple choice vragen omdat ze multiple choice vragen zien als een beoordeling van kennis en lagere orde vaardigheden. Bij essay vragen passen studenten eerder een diepgaande manier van leren toe omdat ze essay vragen zien als een beoordeling van hogere orde vaardigheden. Zij adviseert om essay vragen te gebruiken als docenten studenten willen opleiden tot analytische en kritische denkers, die in staat moeten zijn om kennis in nieuwe situaties toe te passen en om problemen op te lossen. Met essay vragen zijn studenten in de gelegenheid deze hoger orde vaardigheden te demonstreren. **Parmenter** (2009) stelt, op basis van een kwalitatieve survey, dat studenten een voorkeur hebben voor multiple choice vragen omdat ze denken dat deze makkelijker zijn. Ten eerste, er zijn alternatieven die niet geloofwaardig zijn en direct kunnen worden 'weggestreept'. Ten tweede, de gegeven alternatieven maken het gemakkelijk voor de student om het juiste antwoord te

herkennen. Ten derde, is het met name bij wiskundige/cijfermatige vragen mogelijk om terug te redeneren vanuit de gegeven alternatieven. De voorkeur wordt sterker als studenten zich slecht hebben voorbereid. Als ze beter zijn voorbereid neigt hun voorkeur naar een essay toets, als een betere en meer valide beoordeling van hun leerprestaties. De waardering van multiple choice vragen neemt af bij het hogerejaars studenten. Parmenter constateert een strategische aanpak van studenten die hij als volgt omschrijft:

“Any given student would generally apply a mix of surface and deep strategies, with most students having a bias toward one or another and some being adept at adjusting their choice of strategy in reaction to the characteristics of a particular class or assignment. Some student, unfortunately, apply little of either type, seeming unwilling or unable to evaluate their own largely strategy-free behavior or to develop the study skills necessary to become more strategic” (2009, p.60).

Stanger-Hall (2012) vergeleek twee groepen studenten, waarbij de ene groep een toets kreeg met multiple choice vragen, en de andere groep een toets met multiple choice vragen en *constructed response* vragen. Het multiple choice deel bestond in beide toetsen uit 25% vragen gericht op hogere orde vaardigheden. De studenten waren vooraf bekend met de toetsvorm en de aard van de vragen. De studenten die naast multiple choice ook *constructed response* vragen kregen pasten een andere leerstrategie toe en presteerden beter dan de groep met alleen multiple choice vragen. Stanger-Hall concludeert dat het alleen toepassen van multiple choice vragen het kritisch nadenken ontmoedigt. Opvallend gegeven: studenten in beide groepen besteedden evenveel tijd aan het studieonderdeel.

5.3 Validiteit en betrouwbaarheid

In veel van de onderzoeken zijn resultaten op multiple choice toetsen vergeleken met resultaten van andere toetsvormen zoals essay toets en *constructed response test* (aanvulvragen, korte open vragen) met als onderzoeksvraag de validiteit en betrouwbaarheid van multiple choice toetsen. In lijn met de *evidence* ten aanzien van oppervlakkig en diepgaand leren leveren de onderzoeken geen eenduidig beeld op. Een aantal onderzoekers constateert geen verschil in prestaties van studenten tussen multiple choice vragen en andere vraagvormen; (Draaijer & Van den Bos, 2009; Coderre, Harasym, Mandin & Fick, 2004, Walstad & Becker, 1994. O’Neill, 2001). Andere onderzoekers vinden wel verschil (Anbar 1991, Chan & Kennedy, 2002, Becker & Johnston, 1999, Wooten, Cool, Prather & Tanner, 2014) of vinden verschillende effecten, afhankelijk van kenmerken van de student (Pepple, Young, Carroll, 2010).

Draaijer en Van den Bos (2009) vergeleken de slaagpercentages bij een digitale toets tussen traditionele multiple choice vragen en alternatieve vraagvormen (*drag-and-drop, matching, multiple response*). De overall prestaties op een toets met alleen traditionele vragen en een toets, die een mix is van traditionele vraagvormen en alternatieve vraagvormen, leverden vergelijkbare scores op. Echter, de scores op de alternatieve vraagvormen verschilden per vraagvorm. Studenten reageerden positief op de alternatieve vraagvormen; zij voelden zich gedwongen om beter na te denken. **Coderre, Harasym, Mandin en Fick** (2004) onderzochten het verschil tussen multiple choice vragen met vijf antwoord alternatieven en *extended matching* vragen met tien of meer alternatieven op de probleemoplossingsvaardigheden van studenten. De stam, in de vorm van een korte casus, was bij beide vraagvormen identiek. Door studenten hard op te laten redeneren bij het beantwoorden van de vragen constateerden de onderzoekers dat beide vraagvormen in gelijke mate in zijn staat om probleemoplossingsvaardigheden te toetsen. Het aantal alternatieven heeft geen significant

effect op psychometrische aspecten of op de probleemoplossingsstrategie die werd gehanteerd. Ze constateren dat zorgvuldige formulering van de vraag essentieel is om hogere orde vaardigheden te meten; het aantal alternatieven is veel minder van belang.

Anbar (1991) onderzocht het gebruik van multiple choice en aanvulvragen. Hij constateert dat multiple choice vragen een minder betrouwbare meting bieden van kennis dan open einde vragen. Hij vond een significant positieve correlatie tussen multiple choice vragen en aanvulvragen voor de overall lager scorende studenten en een negatieve correlatie tussen multiple choice vragen en open einde vragen voor de overall hoger scorende studenten. Hij leidt hieruit af dat aanvulvragen kennis op een andere manier meten dan multiple choice vragen. Zijn conclusie: competente studenten scoren slechter op multiple choice vragen omdat ze teveel doordenken. **Chan en Kennedy** (2002) onderzochten wat het verschil is in presteren bij multiple choice vragen en *constructed response* vragen. In beide gevallen was de stam van de vraag identiek. Bij de multiple choice vragen kozen de studenten uit vier alternatieven; bij *constructed response* moesten de studenten zelf het antwoord geven. Op *constructed response* vragen, waarbij de vraag weinig context bood om het antwoord af te leiden, scoorden studenten slechter dan bij dezelfde multiple choice vraag.

Pepple, Young, Carroll (2010) voerden een vergelijkbaar onderzoek uit en komen tot andere resultaten. Zij onderzochten de correlatie tussen toetsresultaten op multiple choice vragen en essay vragen in een afsluitende toets. Bij de meeste studenten is er een sterke correlatie tussen toetsresultaten op multiple choice vragen en essay vragen, wat erop wijst dat de prestaties van de studenten niet afhangen van de toetsvorm. Bij studenten aan de uiteinden van het spectrum (hele goede, slechte) constateren de onderzoekers geen correlatie tussen beide toetsvormen: de prestaties in één van de toetsvormen had een sterke invloed op het uiteindelijke cijfer. Voor de goede studenten verklaren ze dit door het feit dat deze studenten heel sterk zijn in òf multiple choice vragen (goed geheugen) òf in essay vragen (analytische vaardigheden). Overall presteerden de studenten beter bij multiple choice vragen in vergelijking met essay vragen. De studenten die niet slaagden waren slecht in beide toetsvormen.

Johnston en Ambusaidi (2000) wijzen erop dat alleen het nakijken van multiple choice vragen een objectieve activiteit is. De keuze voor het soort vragen, de formulering van de alternatieven, de verdeling over de studiestof zijn subjectieve activiteiten. Zelfs de volgorde van de vragen en de volgorde van de alternatieven kan invloed hebben op de prestaties. Uit hun onderzoek blijkt dat het percentage studenten dat het goede antwoord kiest niet stabiel is als de volgorde van de alternatieven verschilt. Dit geldt ook bij een andere volgorde van de vragen.

5.4 Effecten

Multiple choice vragen hebben verschillende effecten. De meest bekende zijn: *testing effect* en *cueing effect*. Het *testing effect* houdt in dat het doen van een toets een positief effect heeft op de prestaties op een latere test, omdat er sprake is van herhaling en herkenning. Dit kan echter een onbedoeld negatief effect hebben. **Roediger & Marsh** (2005) concluderen dat het testing effect negatief kan uitpakken omdat studenten bij multiple choice vragen worden blootgesteld aan foute antwoorden (misinformatie). Hoe meer alternatieven, hoe groter de kans op misinformatie. Uit hun onderzoek blijkt dat het bloot gesteld worden aan foute antwoorden het geheugen kan beïnvloeden, zelfs als de student bij de eerdere toets het goede antwoord heeft gegeven.

Bij het *cueing-effect* bevatten de vraag of de alternatieven aanwijzingen voor het juiste antwoord. Dit sluit aan bij de eerder genoemde perceptie van studenten dat multiple choice vragen makkelijker zijn dan open vragen.

Daarnaast kan veelvuldige inzet van multiple choice leiden tot *test-wiseness* bij studenten. Ze passen strategieën toe bij het beantwoorden van de vragen, zoals het wegstrepen van de twee minst plausibele alternatieven bij vierkeuzevragen. Hierdoor reduceren ze vierkeuze vragen tot *true/false* vragen. Dit vergroot de gokkans en heeft invloed op de uiteindelijke prestatie op de toets (Simkin, 2005).

Er kan ook sprake zijn van *bias* ten aanzien van sekse - mannen zijn beter in multiple choice vragen dan vrouwen - en taalachtergrond - niet-native speakers hebben meer moeite met multiple choice vragen dan native speakers. Uit de literatuur komt geen bewijs naar voren dat er verschil is in prestaties tussen mannen en vrouwen (Becker & Johnston, 1999; Chan & Kennedy, 2002). **Becker en Johnston** (1999) vonden wel een verschil tussen presteren op essay toetsen tussen mannen en vrouwen; vrouwen presteren beter. Er is wel beperkt bewijs dat niet-native speakers slechter presteren op multiple choice vragen dan native speakers in vergelijking met open vragen.

5.5 Leerfunctie

Bij multiple choice vragen als summatieve toets wordt het leereffect over het algemeen als laag beoordeeld. **Nicol** (2007) geeft aan dat multiple choice vragen weinig ruimte geven voor gepersonaliseerde feedback. Ook geven multiple choice vragen weinig ruimte om studenten actief te betrekken bij het bepalen van de doelen en het verhelderen van de beoordelingscriteria. Dat maakt dat het huidige gebruik van multiple choice vragen niet of nauwelijks bijdraagt aan het ontwikkelen van het vermogen van studenten om zelfgestuurd te leren. Hij pleit voor een didactisch *framework* om er voor te zorgen dat docenten optimaal gebruik maken van multiple choice vragen. Aan de hand van de zeven principes van feedback (Nicol & Macfarlane-Dick, 2006) licht hij toe hoe docenten multiple choice vragen effectiever kunnen inzetten (zie paragraaf 3, Tips voor docenten). Ook uit andere onderzoeken blijkt dat het leereffect van multiple choice met relatief eenvoudig interventies kan worden verbeterd. **Butler en Roediger** (2008) constateerden dat feedback op een tussentijdse toets leidt tot betere prestaties op de afsluitende toets. Studenten presteren beter op items in de afsluitende toets, die een tussentijdse toets waren opgenomen, in vergelijking met 'nieuwe' items (testing effect). Bij geen feedback waren de prestaties op de afsluitende toets vergelijkbaar met de tussentijds toets. **Grühn en Cheng** (2014) lieten studenten zelf hun multiple choice vragen na afname corrigeren. In vergelijking met een controlegroep presteren studenten die zelf hun multiple choice toets corrigeerden beter op de afsluitende toets. **Merrel, Schwartz, en Webb** (2015) gingen nog een stap verder en lieten studenten zichzelf corrigeren tijdens de toetsafname. Studenten kregen direct feedback op multiple choice vragen met vijf alternatieven. Ze waren in de gelegenheid om een tweede antwoord te geven. Het percentage goede antwoorden bij tweede gelegenheid, na een eerste foutief antwoord, was significant hoger dan de gokkans. De onderzoekers constateren hieruit dat studenten leren van verkeerde antwoorden.

Gebruikte literatuur

- Anbar, M. (1991). Comparing assessments of students' knowledge by computerized open-ended and multiple-choice tests. *Academic Medicine*, 66(7), 420-422.
- Becker, W.E., Johnston, C. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *The Economic Record*, 75(231), 348-357.
- Biemond, I., Bos-Horstink, M., Soeting, J., Sugito, A., Uil-Hogerwaard, W. (2015). *Hoe maak ik goede toetsen met gesloten en open vragen?* Teelen B.V.
- Buckles, S., Siegfried, J.J. (2006). Using multiple-choice questions to evaluate in-depth learning of economics. *The Journal of Economic Education*, 37(1), 48-57.
- Bulter, A.C., Roediger, H.L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604-616.
- Coderre, S.P., Harasym, P., Mandin, H., Fick, G. (2004). The impact of two multiple choice question formats on the problem-solving strategies used by novices and experts, *BMC Medical Education*, 4(23).
- Chan, N., Kennedy, P.E. (2002). Are multiple choice exams easier for economics students? A comparison of multiple-choice and 'equivalent' constructed-response exam questions. *Southern Economic Journal*, 68(4), 957-971.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Draaijer, S., Van den Bos, G.C. (2009). Tentamen met de computer: een vergelijking van meerkeuzevragen en alternatieve vraagvormen. *Tijdschrift voor Medisch Onderwijs*, 28(1), 13-21.
- Dufresne, R.J., Leonard, J.L., Gerace, J.G. (2002). Making sense to students' answers on multiple choice questions. *The Physics Teacher*, 40, 174-180.
- ECABO (2012), *Schrijfwijzer gesloten vragen*, SURF.
- ECABO (2012), *Schrijfwijzer toetstermen*, SURF.
- Elton, L., & Laurillard, D. (1979). Trends in research on student learning. *Studies in Higher Education*, 4(1), 87-102.
- Gibbs, G. (2010). *Using assessment to support student learning*. Leeds Met Press.
- Grühn, D., Cheng, Y. (2014). A self-correcting approach to multiple-choice exams improves students' learning. *Teaching of Psychology*, 41(4), 335-339.
- Johnstone, A.H., Ambusaidi, A. (2000). Fixed response: what are we testing? *Chemistry Education: Research and practice in Europe*, 1(3), 323-238.
- Merrel, J.D., Schwartz, P.M., Webb, J.A. (2015). Multiple choice testing using immediate feedback – assessment technique (IFAT) forms: assessing learning from mistakes. *Higher Education Studies*, 5(5), 50-55.
- Nicol, D., Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 198-218.
- Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education*, 31(1), 53-64.
- Nieweg, M. (2014). De mc-toets: een krasse grijsaard van honderd jaar. *Examens*, (3), 21.
- O'Neill, P.B. (2001). Essay versus multiple choice exams; an experiment in the principles of macroeconomics course. *The American Economist*, 45(1), 62-70.
- Palmer, E.J., Devitt, P.G. (2007). Assessment of higher order cognitive skills in undergraduate education; modified essay or multiple choice questions? *BMC Medical Education*, 7:49.
- Parmenter, D.A. (2009). Essay versus multiple choice: student preferences and the underlying rationale with implications for test construction. *Academy of Educational Leadership Journal*, 13(2).
- Pepple, D.J., Young, L.E., Carroll, R.G. (2010). A comparison of student performance in multiple choice and long essay questions in the MBBS stage I physiology examination at the University of the West Indies (Mona Campus). *Advances in Physiology Education*, 34, 86-89.
- Roediger, H.L., Marsh, E.J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology*, 31(5), 1155-1159.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches; multiple choice questions examination versus assignment essay. *Higher Education*, 35(4), 453-472.
- Simkin, M.G., Kuechler, W.L. (2005). Multiple choice tests and student understanding; what is the connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73-97.

- Stanger-Hall, K.F. (2012). Multiple choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE-Life Sciences Education*, 11, 294-306.
- Tractenberg, R.E., Gustha, M.M., Mulroney, S.E., Weissinger, P.A. (2013). Multiple choice questions can be designed or revised to challenge learners' critical thinking. *Advances in Health Science Education*. 18, 945-961.
- Visser, D.Y., Bootsma, A.L., Ten Cate, Th.J. (2010). Kennis, inzicht en toepassing: welke typen toetsvragen passen hierbij? *Tijdschrift voor Medisch Onderwijs*, 29(6), 337-34.
- Walstad, W.B., Becker, W.E. (1994). Achievement differences on multiple choice and essay tests in economics. *The American Economic Review*, 84(2), 193-196.
- Wooten, M.M., Cool, A.M., Prather, E., Tanner, K.D. (2014). Comparison of performance on multiple-choice questions and open-ended questions in an introductory astronomy laboratory. *Physics Education Research*, 10(2), 1-22.

Bijlage 1: Overzicht onderzoeksartikelen

Selectiecriteria literatuur survey: na 1990, uit een peer reviewed tijdschrift, gericht op hoger onderwijs.

Onderzoeksvraag: vergelijking multiple choice (mc) vragen met andere toetsvormen

Eerste auteur*	jaartal	Land	Type onderzoek	Onderzoeksvraag	Uitkomsten
Anbar	1991	USA	Case study; vergelijking mc vragen- open einde vragen	Meten mc vragen en open einde vragen op dezelfde wijze en in dezelfde mate kennisontwikkeling?	-Een significant positieve correlatie tussen de mc vragen en open eind vragen voor de overall lager scorende studenten; een negatieve correlatie tussen mc vragen en open einde vragen voor de overall hoger scorende studenten. -Suggesties dat open einde vragen kennis op een andere manier wordt gemeten dan bij mc vragen. -Competente studenten scoren slechter op mc vragen; ze denken teveel door. -Mc vragen bieden een minder betrouwbare meting van kennis dan interactieve open eind vragen.
Walstad	1994	USA	Case study; vergelijking mc vragen en essay vragen	Testen mc vragen en essay over dezelfde studiestof overeenkomstig begrip van micro en macro-economie?	-Mc vragen en essay vragen meten een overeenkomstig begrip van economische onderwerpen. -Essay vragen, gegeven de hogere kosten voor beoordelen, inzetten bij een duidelijke toegevoegde waarde.
Scouller	1998	AUS	Case study; vergelijking mc vragen en essay, perceptie studenten	Hoe verhoudt de voorkeur voor een bepaalde toetsvorm van de studenten zich tot de leerstrategie die hij/zij toepast en de perceptie van wat beoordeelt wordt?	-Studenten passen oppervlakkig leren toe bij mc vragen en beschouwen mc vragen als beoordeling van kennis, lagere orde vaardigheden. -Bij essays passen studenten eerder een diepgaande manier van leren toe en zien essays als een beoordeling van hogere orde vaardigheden.
Becker	1999	AUS	Case study; vergelijking mc vragen en essay vragen	Testen mc vragen en essay over dezelfde studiestof overeenkomstig begrip van micro en macro-economie?	-De score op een mc test is geen goede voorspeller voor score op essay vragen en andersom. -Vrouwelijke studenten scoren beter op essay vragen dan mannelijke studenten; voor mc vragen zijn de scores vergelijkbaar.
O'Neill	2001	USA	Experiment, verschil in presteren tussen mc vragen en constructed response	Wat is de impact van toetsvorm op prestaties van studenten?	-Toetsvorm – mc vragen of constructed response - heeft geen invloed op gemiddelde prestaties van beide student groepen. -Toetsvorm heeft wel invloed op de TUCE score (gestandaardiseerde, nationale test).
Chan	2002	CAN	Experiment; vergelijking mc en constructed response (cr) vragen, perceptie studenten	Scoren studenten gemiddeld beter of slechter op mc vragen ten opzicht van cr vragen?	-Op cr vragen waar studenten een kort antwoord moeten formuleren, waarbij de vraag weinig context biedt om het antwoord af te leiden, scoren studenten slechter dan bij dezelfde mc vraag. -Er is geen verschil tussen mannen en vrouwen of goede versus slechtere studenten.

Coderre	2004	CAN	Case study; vergelijking vijf alternatieven vragen en matching vragen	Zijn deze twee vraagvormen in staat om probleem oplossings vaardigheden te toetsen?	-Beide vraag vormen zijn in gelijke mate in staat om probleem oplossings vaardigheden te toetsen. -Zorgvuldige formulering van de vraag is essentieel om hogere orde vaardigheden te meten, aantal alternatieven is veel minder van belang. -Het aantal alternatieven heeft geen significant effect op psychometrische aspecten of de probleem oplossings strategie die werden gehanteerd.
Palmer	2007	AUS	Inhoudelijke analyse: vergelijking mc vragen en modified essay questions	Zijn modified essay questions effectiever in het beoordelen van hogere orde vaardigheden	-Goed geconstrueerde mc vragen kunnen voldoen aan het cognitieve niveau dat <i>modified essay questions</i> vragen beogen. -Ook bij <i>modified essay questions</i> is constructie de achilleshiel; teveel gericht op reproductie.
Draaijer	2009	NL	Case study; vergelijking traditionele mc vragen en alternatieve gesloten vraagvormen	Verandert het slaagpercentage door alternatieve vraagvormen. Zo ja, hoe?	-Alternatieve vraagvormen leveren scores op in dezelfde orde grootte als meerkeuzevragen. -Alternatieve vraagvormen resulteren in verschillende slaagpercentages (experiment is uitgevoerd bij herkansers).
Parmenter	2009	USA	Case study; vergelijking mc vragen en essay; perceptie studenten	Welke toetsvorm heeft de voorkeur bij studenten en waar is de rationale hiërachter?	Studenten hebben een voorkeur voor mc vragen omdat ze denken dat deze makkelijker zijn. De voorkeur wordt sterker als studenten zich slecht hebben voorbereid. Als ze beter zijn voorbereid neigt hun voorkeur naar essay, als een betere en meer valide beoordeling van hun leerprestaties. De waardering van mc vragen neemt af bij het hogere jaars studenten.
Peple	2010	JAM/USA	Case study: vergelijking mc en essay	Is er een correlatie tussen prestaties van studenten bij mc en essay vragen?	-Studenten presteren beter bij mc vragen in vergelijking met essay vragen. -Bij de meeste studenten is er een sterke correlatie tussen toetsresultaten op mc vragen en essay vragen, wat erop wijst dat prestaties van studenten niet afhangen van de toetsvorm. -Bij studenten aan de uiteinden van het spectrum (hele goede, slechte) is geen correlatie geconstateerd, dit wijst er op dat hun prestaties in een van de toetsvormen een sterke invloed heeft op de uiteindelijke resultaat. -De studenten die niet slaagden waren slecht in beide toetsvormen.
Stanger-Hall	2012	USA	Experiment; vergelijking mc en constructed reponse (cr) vragen	Leidt het aanpassen van een toets van alleen mc naar combinatie van mc en cr, tot verandering in de leeruitkomsten en een verbetering van kritisch nadenken?	-Een toets met alleen mc vragen ontmoedigt het kritisch nadenken. -Perceptie van de te verwachten toets heeft effect op de leerstrategie en prestaties.
Wooten	2014	USA	Case study; vergelijk mc vragen en open-ended questions	Is het presteren op mc vragen vergelijkbaar met het presteren op open vragen?	-Studenten presteren beter op de mc vragen in vergelijking met aanvullende de open eind vragen, waarbij studenten werden gevraagd hun mc antwoord toe te lichten. -Het kiezen van het juiste antwoord op een mc vraag kan een overschatting zijn van begrip van de student

* zie referenties voor de volledige gegevens

Onderzoeksvraag: Validiteit en betrouwbaarheid multiple choice vragen

Eerste auteur*	jaartal	Land	Type onderzoek	Onderzoeksvraag	Uitkomsten
Johnstone	2000	UK	Inhoudelijke analyse	Is mc een valide en betrouwbaar toetsinstrument?	-Het percentage studenten dat het goede antwoord kiest is niet stabiel als de volgorde van de alternatieven verschilt. -Dit geldt ook bij een andere volgorde van de vragen.
Dufresne	2002	USA	Case study; validiteit van de mc vragen	Is de juiste beantwoording van een mc vraag een goede indication voor begrip van de student?	-Het goede antwoord op één vraag over bepaalde materie kan in hoge mate kan berusten op gokkans, intuïtie of onvolledige kennis. -Bij het hanteren van één vraag over bepaalde materie kan het juiste antwoord ten onrechte de indruk geven dat de student de materie begrijpt.
Simkin	2005	USA	Inhoudelijke analyse	Meten mc en cr vragen hetzelfde niveau van begrijpen; In welke mate is mc voorstellend voor presteren op cr?	-Mc vragen zijn te construeren voor de eerste drie niveaus van de taxonomie van Bloom (kennis, begrip, toepassen). -Het vierde niveau: analyse laat zich moeilijk vatten in mc vragen omdat dit gevat moet worden in een juist antwoord. -Mc vragen gericht toe toepassen zijn lastig te construeren, cr vragen zijn hiervoor geschikter.
Roediger	2005	USA	Testing effect bij mc vragen	Heeft een eerdere mc test een positief testing effect op de final test? Leidt meer alternatieven tot een kleiner testing effect?	-Eerdere mc kennis test leidt tot een hogere score bij de final test in vergelijking met alleen een final mc test. -Het lezen van foute afleiders vermindert het positieve testing effect kan leiden tot foute antwoorden bij de final test. -Meer alternatieven levert een kleiner testing effect op. -Mc vragen kunnen leiden tot het foute kennis.
Buckles	2006	USA	Inhoudelijke analyse	Zijn mc vragen in staat om economisch redeneren en analyseren te toetsen?	-Met het gebruik van multiple choice vragen kan diepgaand leren getoetst worden. -Het formuleren van vragen op een hoger niveau van Bloom vereist wel de nodige vaardigheid en tijd.
Tractenberg	2013	USA	Inhoudelijke analyse: complexiteit en moeilijkheidsvraag	Kan cognitieve complexiteit van een mc vraag los worden gezien van moeilijkheidsgraad?	-Cognitieve complexiteit en moeilijkheidsgraad van mc vragen zijn onafhankelijke dimensies -Training van toetsexperts in cognitieve complexiteit kan helpen om de hogere niveaus van Bloom in mc vragen tot uiting te laten komen.

* zie referenties voor de volledige gegevens

Onderzoeksvraag: Leeropbrengst en feedback bij multiple choice vragen

Eerste auteur*	jaartal	Land	Type onderzoek	Onderzoeksvraag	Uitkomsten
Nicol	2007	UK	Inhoudelijke analyse	Hoe kan de leerfunctie van mc vragen worden verhoogd met gebruikmaking van feedback?	-De toepassing van <i>seven principles of good feedback practice</i> op multiple choice vragen bieden mogelijkheden om de leerfunctie van multiple choice vragen te verhogen.
Butler	2008	USA	Experiment: effect van feedback bij mc vragen	Kan feedback het negatieve effect van mc toets verminderen en de positieve effect versterken?	-Studenten presteren beter op items in de final test, die een tussentijdse toets waren opgenomen, in vergelijking met 'nieuwe' items (testing effect). -Bij geen feedback waren de prestaties op de final test vergelijkbaar met de tussentijdstoets. -Feedback op de tussentijdse test verhoogde het aantal goede antwoorden op de final test, in vergelijking met de tussentijdse test.
Grühn	2014	USA	Experiment: leereffect van het zelf corrigeren van mc toets	Heeft het zelf corrigeren van een mc toets een positief leereffect?	-Studenten die zelf hun mc toets corrigeerden presteerden beter op de afsluitende toets.
Merrel	2015	USA	Case study; effect van directe feedback bij mc vragen	Scoren studenten beter als ze tijdens de toets de gelegenheid hebben om hun antwoord te corrigeren?	-Het percentage goede antwoorden bij correctie van een fout antwoord is significant hoger dan de gokkans: studenten leren van verkeerde antwoorden.

* zie referenties voor de volledige gegevens