Opgesteld door: Rena Punt (kennismakelaar)
Vraagsteller: Extern lid examencommissie lerarenopleiding
Referentie: Kennisrotonde (2020). *Welke invloed heeft het hergebruiken van vragen in kennistoetsen op de kwaliteit van de toets binnen de lerarenopleiding?* (KR. 953). Den Haag: Kennisrotonde.

22 juli 2020

Vraag
Welke invloed heeft het hergebruiken van vragen in kennistoetsen op de kwaliteit van de toets binnen de lerarenopleiding?

Kort antwoord
Als een kennistoets voor een deel bestaat uit eerder gebruikte vragen die al bij studenten bekend (kunnen) zijn, leidt dit niet automatisch tot een daling van de kwaliteit (betrouwbaarheid en validiteit) van de toets. De vier aangetroffen internationale onderzoekspublicaties beschrijven alle een situatie waarin het hergebruiken van multiple-choicevragen in een kennistoets binnen het hoger onderwijs weinig tot géén effect had op zaken als slagingspercentage en moeilijkheidsgraad van de toets(vragen). Er is overigens geen onderzoek bekend naar hergebruik van open vragen, of situaties waarin er géén sprake is van een itembank. Evenmin is bekend wat de invloed is van andere toetskenmerken en studentfactoren op de kwaliteit van de toets(vragen) bij hergebruik van vragen.

Toelichting antwoord
Het doel van een kennistoets[1] is beoordelen in hoeverre een student over de vereiste kennis beschikt: feiten weten te reproduceren, inzicht laten zien en/of kennis weten toe te passen. Soms kiezen docenten er voor om toetsvragen of toetsen ongewijzigd te hergebruiken. Ze nemen bijvoorbeeld eerdere oefenvragen op in een toets of gebruiken vragen nog eens bij een herkansing, bij parallelle studentengroepen of bij volgende cohorten. De vragen en antwoorden kunnen op den duur, bedoeld of onbedoeld, bekend raken onder studenten die de kennistoets nog(maals) moeten afleggen.

Het ongewijzigd hergebruiken van vragen in een kennistoets kan voordelen hebben, maar vormt ook een mogelijke bedreiging voor de kwaliteit van de toets(vragen). Onderzoeker en toetsexperts noemen – overigens niet op basis van eigen of ander

[1] Dit antwoord op een vraag aan de Kennisrotonde heeft enkel betrekking op kennistoetsen. Andere toetsvormen zijn buiten beschouwing gelaten.

LAAN VAN NIEUW OOST INDIË 300 2593 CE DEN HAAG
POSTADRES POSTBUS 93461 2509 AL DEN HAAG
T 070 349 44 58 W WWW.KENNISROTONDE.NL

empirisch onderzoek – de onderstaande mogelijke voordelen en risico's (zie bijvoorbeeld Gehringer, 2004; Molkenboer & Kemper, 2017; Van Berkel et al., 2014; Wagner-Menghin et al., 2013; Yang, 2012):

Mogelijke voordelen	Mogelijke risico's
Aangetoonde kwaliteit: van reeds gebruikte toetsvragen kan men op basis van psychometrische analyse vaststellen dat deze betrouwbaar en valide zijn voor het betreffende doel.	Hergebruik kan ten koste gaan van de validiteit omdat de vragen (als ze bekend zijn bij studenten) makkelijker kunnen worden om te beantwoorden.
Gelijkwaardigheid en vergelijkbaarheid: het gebruik van dezelfde toetsvragen voor verschillende studenten(groepen) kan bijdragen aan gelijkwaardigheid van de toetsen tussen de groepen en de onderlinge vergelijkbaarheid van de toetsresultaten.	Bovenstaande kan een oneerlijk voordeel opleveren voor studenten die onbedoeld de hand hebben weten te leggen op eerdere toetsvragen. Daarmee samenhangend kan het onderscheidende (discriminerende) vermogen van de toets afnemen wat betreft studenten die de kennis wel beheersen en zij die de kennis niet beheersen.
Transparantie: het bekend maken van vragen kan eraan bijdragen dat studenten beter weten wat ze te wachten staat bij de toets.	Hergebruik werkt 'teaching to the test' in de hand, waardoor het leerrendement kan afnemen, met name als studenten alleen leren wat zij verwachten tegen te komen in de toets in plaats van dat ze de gehele leerstof leren.
Als een verzameling van toetsvragen de gehele leerstof bestrijkt, kan het bestuderen van deze gehele vragenset een goede manier zijn om de beoogde (feiten)kennis te leren, en zou hiermee tevens borgen dat alle lesstof aangeboden wordt; als iets niet in het onderwijs aan de orde is gekomen, dan gebeurt dit alsnog via de toetsvragen.	
Hergebruik scheelt ontwikkeltijd en –kosten.	

Hergebruik vragen leidt niet automatisch tot kwaliteitsdaling.
Uit het (beperkt aanwezige) onderzoek naar ongewijzigd hergebruik van vragen in kennistoetsen in het hoger onderwijs komt het beeld naar voren dat de validiteit en betrouwbaarheid van de toets en/of de toetsvragen onder bepaalde condities niet of nauwelijks wordt beïnvloed wanneer de toets vragen bevat die hergebruikt zijn en bekend zijn bij de studenten.

Wagner-Menghin et al. (2013) concluderen met de nodige voorzichtigheid dat het hergebruiken van een beperkt aantal vragen in een multiple-choice kennistoets (35-45% van het totaal aantal vragen in de toets) de examenkwaliteit niet automatisch beïnvloedt. Zij vonden in hun experimenteel onderzoek onder 671 Oostenrijkse geneeskundestudenten dat toepassingsgerichte multiplechoicevragen bij hergebruik beter gemaakt werden, en dus bij hergebruik een lagere moeilijkheidsgraad hadden. Dat wil

zeggen dat het percentage studenten dat het juiste antwoord gaf steeg. Dit vormt een bedreiging voor de validiteit omdat de vragen mogelijk minder geschikt worden om onderscheid te maken tussen studenten die de leeruitkomsten wél beheersen en zijn die dat niet doen. Echter, bij de reproductievragen bleek er bij hergebruik geen sprake te zijn van een daling van de moeilijkheidsgraad. De hergebruikte vragen werden bij verschillende gelegenheden ongeveer even goed beantwoord. De overallscores van de studenten op de tentamens mét en zonder hergebruikte vragen verschilden niet significant. Zij onderzochten verder of studenten die er voor kozen om later tentamen te doen gemiddeld hoger scoren dan studenten die het tentamen als eerste maken. De hypothese was dat deze studenten strategisch hun tentamenmoment plannen zodat er gelegenheid is tot het uitwisselen van informatie, en dat zij hoger zouden scoren. Ook daarvoor bleek bij dit onderzoek geen bewijs; de studenten die als laatste tentamen deden hadden zelfs de laagste scores. De vraag is hoe het kan dat zij niet beter presteerden, ondanks dat zij gelegenheid hadden om profijt te hebben van het 'lekken' van vragen. De studenten zullen volgens de onderzoekers de eerdere toetsvragen wel besproken en bestudeerd hebben; de moeilijkheidsgraad van de toepassingsvragen daalde immers. Een verklaring zou kunnen zijn dat de studenten die willen profiteren van kennis over de vragen - en dus een later toetsmoment kiezen - de zwakkere studenten zijn. Juist deze studenten zijn mogelijk beperkt in staat om voordeel te behalen uit het (voor)bespreken van vragen.

Ook bij Canadees onderzoek van Wood (2008) maakte het voor de prestaties van de studenten nauwelijks uit of hun hertentamen, een kennistoets, hergebruikte vragen bevatte of niet. 130 studenten geneeskunde deden dit hertentamen: de ene groep kreeg daarbij alleen nieuwe vragen (324 multiple-choicevragen), en de andere groep kreeg 11% tentamenvragen die ook in het eerdere tentamen waren opgenomen. De studenten wisten niet van tevoren dat er vragen zouden worden hergebruikt en tussen de eerste en de tweede afname zat ongeveer 4 maanden. De studenten scoorden gemiddeld beter dan de eerste keer, zonder significant verschil tussen beide groepen.

Yang (2012) beschrijft een situatie waarin het bewust bekend maken van multiple-choicevragen geen significant effect had op de kwaliteit van de toets. De nationale raad voor zorgpersoneel van Korea besloot alle 5548 multiple-choicevragen van de landelijke 'Medical Licensing Examination' te publiceren in het kader van transparantie richting de kandidaten. In totaal legden 20.455 studenten een examen af dat bestond uit 400 tot 540 vragen, samengesteld uit de set van 5548 vragen. De ene groep deed examen voorafgaand aan publicatie, de andere groep daarna. Er werden geen significante verschillen gevonden in slagingspercentage, prestaties van de studenten of de moeilijkheidsgraad van de toets. Het vrijgeven van alle 5548 vragen die tezamen het gehele kennisdomein van de toets bestrijken, had dus geen effect op de kwaliteit van de toets. De grote omvang van de totale vragenset, het aantal vragen in de toets en het feit dat de 5548 vragen de gehele leerstof omvatten zijn in deze situatie mogelijk factoren van belang.

Hertz et al. (2003) onderzochten het effect van hergebruik van vragen in examens voor LSCW (Licensed Clinical Social Worker) en MFT (marriage and family therapist) in de Verenigde Staten. In totaal 2.661 studenten deden in een periode van 6 maanden mee

aan een eerste, tweede of derde examengelegenheid. Het betrof examens met 175 multiple-choicevragen uit een itembank, waarbij iedere kandidaat de vragen in een andere volgorde via de computer kreeg aangeboden. De inhoud van de vragen was veelal toepassingsgericht. De overlap tussen de vragen in de eerste en latere gelegenheden was ongeveer 30%. Er was geen significante toename van het slagingspercentage van het examen of het percentage studenten dat de hergebruikte vragen juist beantwoordde. Dus de studenten leken geen voordeel te halen uit het feit dat er hergebruikte vragen in het examen zaten. Volgens de auteurs draagt goed gebruik maken van een digitaal toetssysteem er aan bij dat er efficiënt gebruik gemaakt kan worden van de itembank (door middel van itemanalyse en randomiseren van examenvragen). De onderzoekers melden in hun publicatie niet wat de omvang van de itembank was.

Tot slot

De vier aangehaalde onderzoeken zijn alle afkomstig uit het buitenland en hebben betrekking op kennistoetsen in hoger onderwijs binnen zorg en welzijn. Er is echter geen reden om aan te nemen dat het land of het kennisdomein van de toets(vragen) verschil zou maken voor het effect van hergebruik van vragen op de kwaliteit van de toets, waardoor deze onderzoeken in die zin ook voor de lerarenopleiding interessant kunnen zijn. Mogelijk beperkend voor de vergelijkbaarheid met de situatie bij de lerarenopleiding is wel dat de onderzoeken niks zeggen over het effect van hergebruik van open vragen of het hergebruik van vragen in een situatie zonder itembank. Verder is onbekend in welke mate overige contextfactoren een rol spelen bij het effect van hergebruik van toetsvragen op de kwaliteit. Denk aan studentfactoren (zoals de mate waarin zij bereid en in de gelegenheid zijn geweest om vragen uit te wisselen en of zij verwachten dat vragen hergebruikt worden) en toetskenmerken zoals het percentage hergebruikte vragen in de toets, de verhouding tussen de bekend geraakte vragen en de lesstof (dekking), het percentage bekend geraakte vragen van de totale vragenset en de wijze waarop de vragen worden voorgelegd in de toets (vaste volgorde of gerandomiseerd). Het is denkbaar dat deze factoren bepalend zijn voor de mate waarin hergebruik de kwaliteit van de toets(vragen) beïnvloedt, maar onderzoek hiernaar ontbreekt.

Meer weten?

Kennisrotonde: (2020) *Op welke manier is oefenen met toetsen in het basisonderwijs effectief voor betere beheersing van de te toetsen lesstof in plaats van teaching to the test?* (KR.836). Den Haag

Geraadpleegde bronnen

Case, S., & Swanson, D. (2002). *Constructing Written Test Questions For the Basic and Clinical Sciences.* National Board of Examiners.

McCabe, D.L., Butterfield, K.D., & Treviño, L.K. (2006). *Academic dishonesty in graduate business programs: prevalence, causes, and proposed action.* *Academy of Management Learning and Education, 5(3),* 294–305.

Gehringer, E. (2004). *Reuse of homework and test questions: when, why, and how to maintain security? Proceedings - Frontiers in Education Conference. Vol. 3*.

Hertz, N. R., Chinn, R.N., & Roberta, N. (2003). *Effects of question exposure for conventional examinations in a continuous testing environment.* Paper presented at the 2003 Annual Meeting of the National Council on Measurement in Education, Chicago.

Krebs, R. (2004). *MC-Prüfungen für die ärztliche Ausbildung. Institut für Medizinische Lehre*, Abteilung für Ausbildungs- und Examensforschung. Bern, Zwitserland.

Molkenboer, H.F.A.M., & Kempers, H.H.M (2017). *Oppassen met het hergebruik van tentamenvragen.* *Examens, 2017 (1),* 7-10.

Teixeira, A. A., & Rocha, M.F. (2010). *Cheating by economics and business undergraduate students: an exploratory international assessment*. *Higher Education, 59(6),* 663–701.

Van Berkel, H.J.M., Bax, A.E., & Joosten-ten Brinke (2014). *Toetsen in het hoger onderwijs* (3de druk). Houten: Bohn Stafleu van Loghum.

Wood, T.J. (2009). *The effect of reused questions on repeat examinees.* *Advances in Health Sciences Education, 14(4),* 465-73.

Wagner-Menghin, M., Preusche, I., & Schmidts, M. (2013). *The Effects of Reusing Written Test Items: A Study Using the Rasch Model.* *ISRN Education, 2013,* 7p.

Yang, E.B., Lee, M.A., & Park, Y.S. (2018). *Effects of test item disclosure on medical licensing examination.* *Advances in Health Sciences Education, 23,* 265–274.

Onderwijssector
Hoger onderwijs

Trefwoorden
Toetsing, examinering, kennistoets