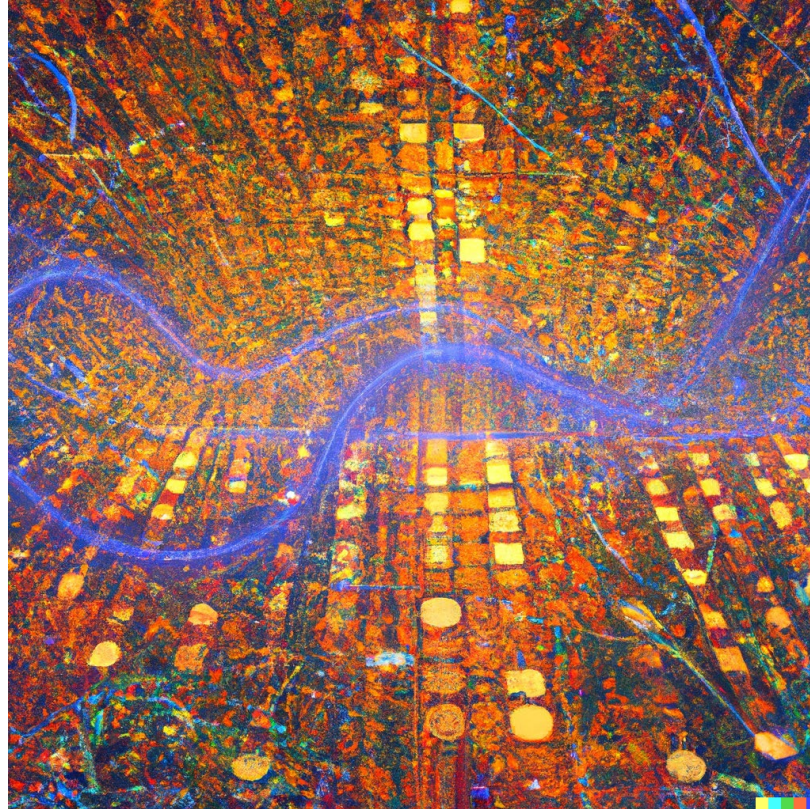




Toetsen voor leren/ programmatisch toetsen

Master class toetsen

Het is complex



Veranderingen in de wereld

- Democratisering van kennis
- Democratisering van transacties/vertrouwen
- Democratisering van doceren
- Democratisering van communicatie
- Democratisering van beschikbaarheid
- Democratisering van intelligentie/cognitie

moderne student 'affordances'

communicatie/communities

collaboraties

creaties

convergentie

Mijn studie

Tutor groepen

Colleges

Practica

Co-schappen

MCQs/OEQs

Mondelinge examens

OSCEs

Essays

Huidige studenten

TBL groepen

(Opgenomen) colleges

Practicum

Co-schappen

MCQs/OEQs

Mondelinge examens

OSCEs

Essays

Portfolio

Mijn studie



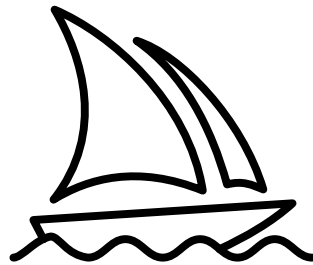
Huidige studenten



Nu ook generative AI



How can I help you today?



DALL-E 3

Verschillende visies

onderwijs

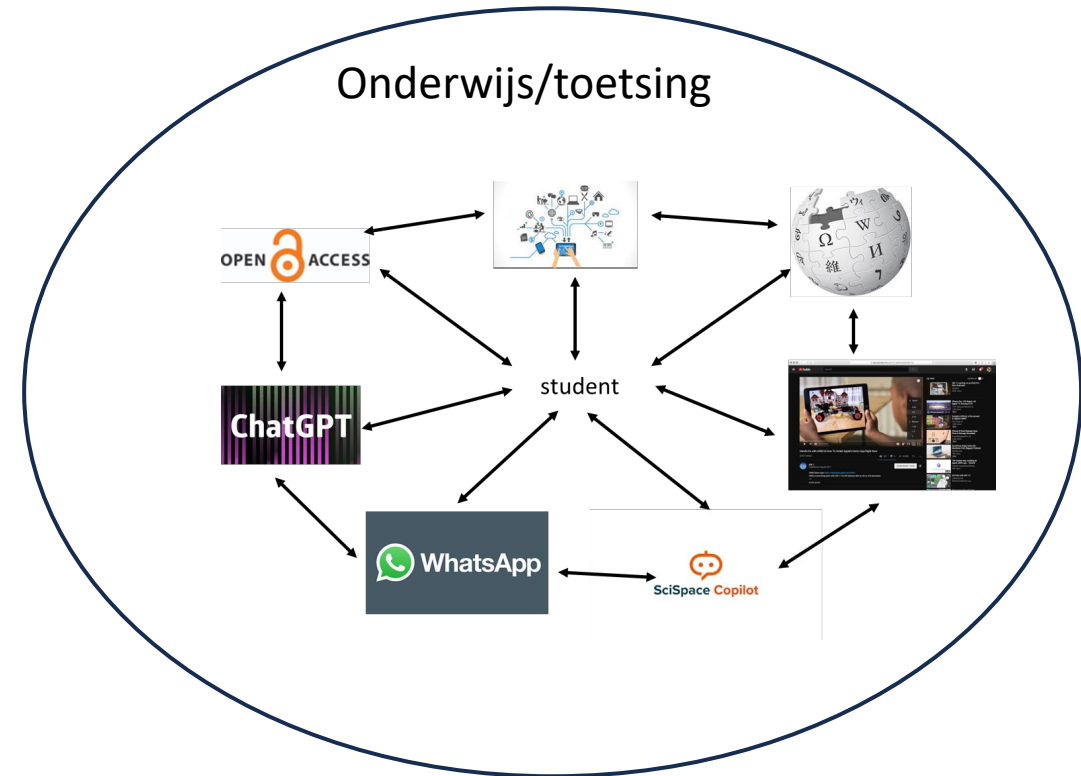


student



toetsing

*Cognitive offloading
cheating*

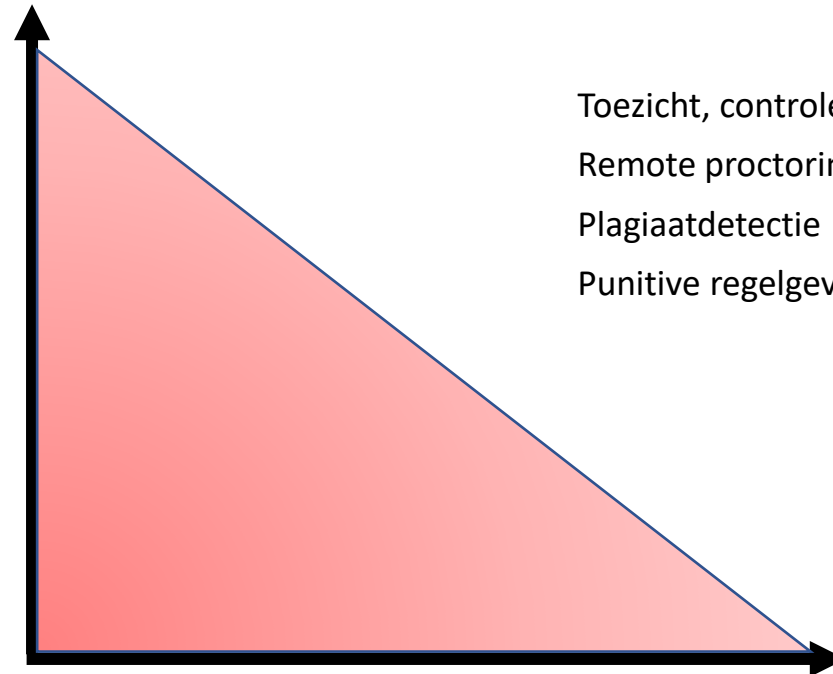


Distributed cognition

Waar gaat het om?

Levenslang leren
Zelfstandig leren
Competentie
Samenwerken
Onzekerheid

onderwijs

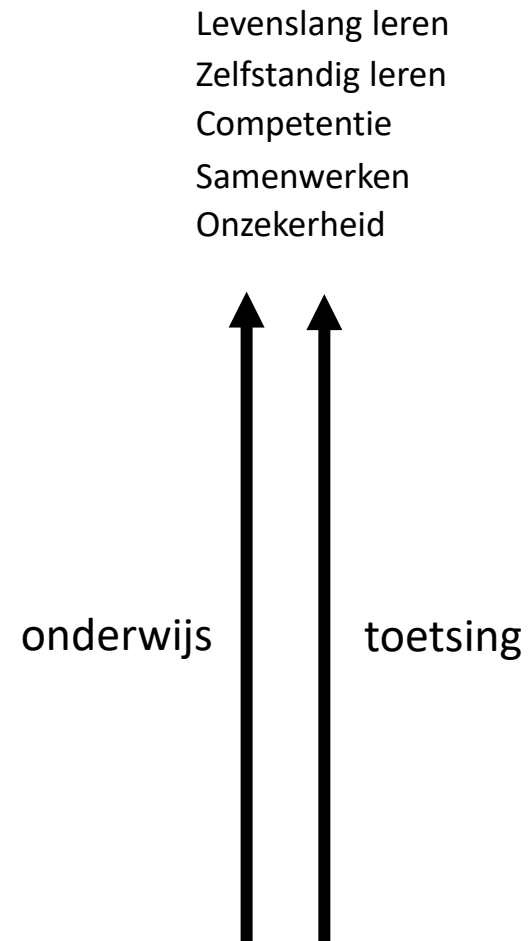


Toezicht, controle
Remote proctoring
Plagiatdetectie
Punitive regelgeving

toetsing

Individuele competentie
Voorgeschreven activiteit
Onderscheiden
Rangordening
Cijfers

Waar gaat het om?



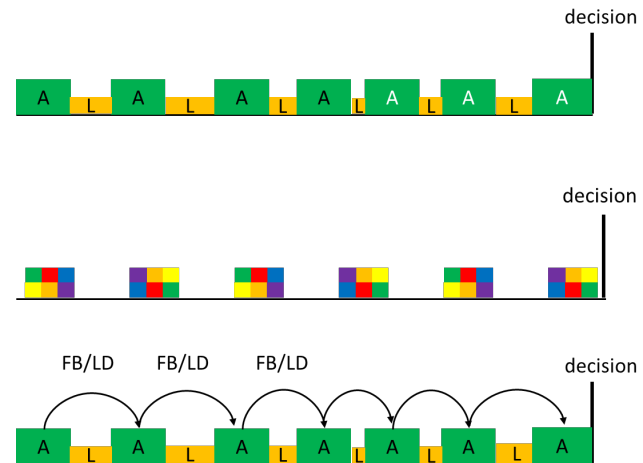
Assessment for learning

- Inhoud:
 - Betekenisvol
 - Actief
 - Samenwerkend

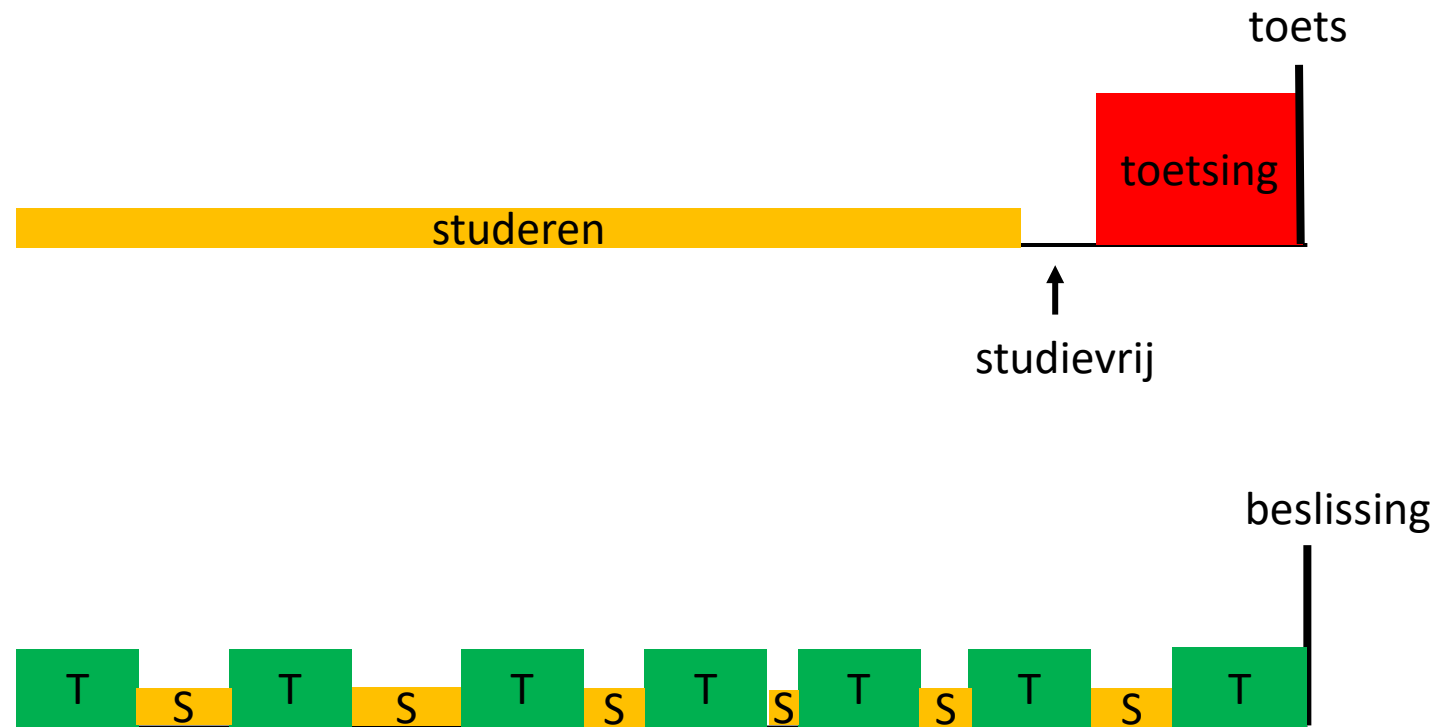
Assessment for learning

- Proces:

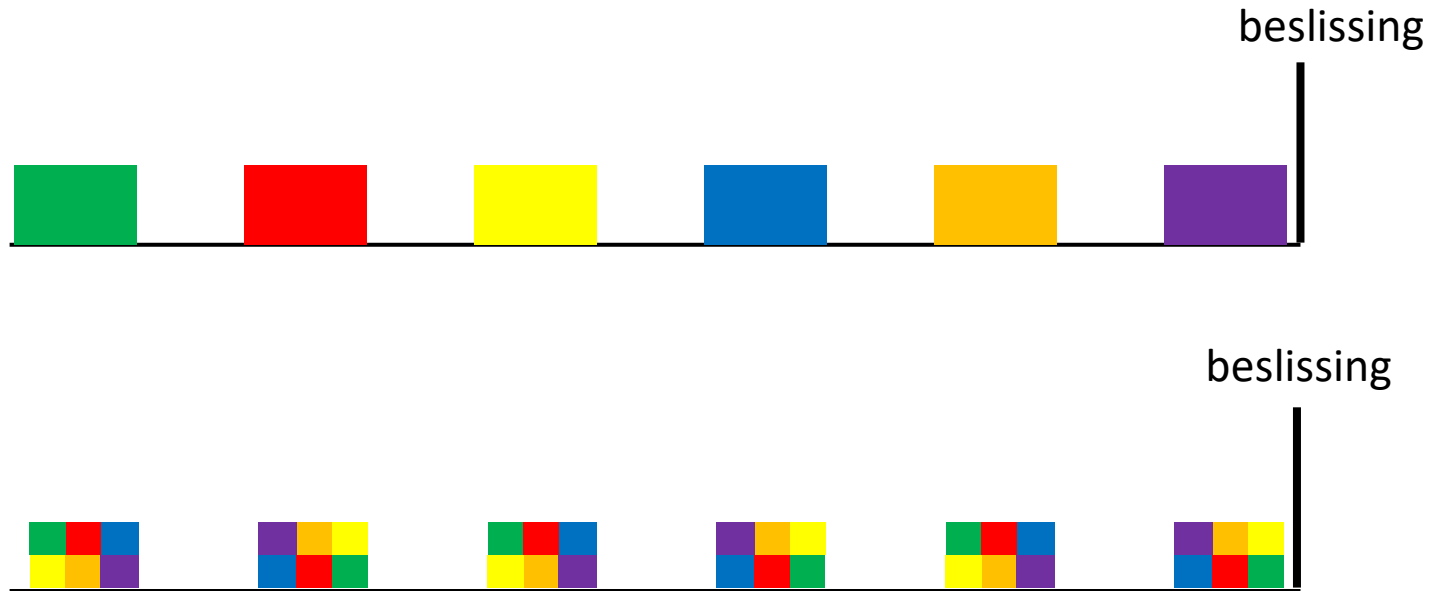
- Distributed
- Interleaved
- Connected



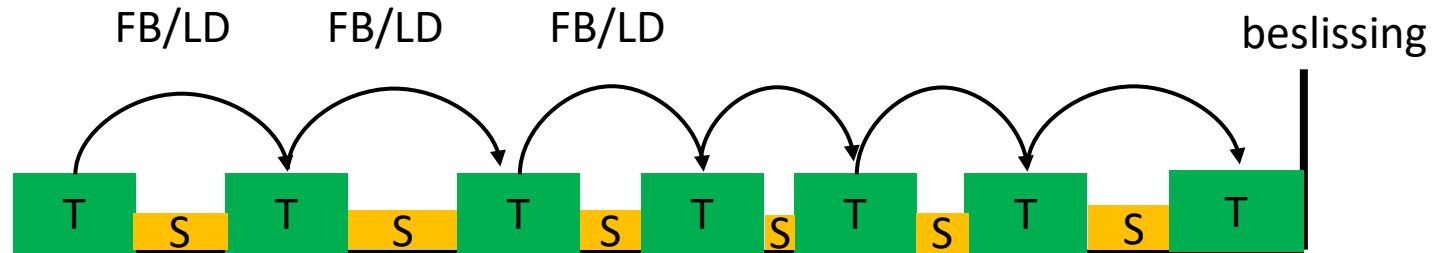
Assessment for learning: distributed



Assessment for learning: interleaved



Assessment for learning: connected



Assessment for learning → niet

Score 1 ← Assessment 1 → *Feedback + LG 1*

Score 2 ← Assessment 2 → *Feedback + LG 2*

Score 3 ← Assessment 3 → *Feedback + LG 3*

Score 4 ← Assessment 4 → *Feedback + LG 4*

Score 5 ← Assessment 5 → *Feedback + LG 5*

Score 6 ← Assessment 6 → *Feedback + LG 6*

Assessment for learning → wel



Waarom programmatisch toetsen

oud model van medische competentie

kennis

vaardigheden

probleemoplossen

attitude

TOETS

TOETS

TOETS

TOETS

TOETS

TOETS

TOETS

TOETS

Competentiemodellen



CanMeds

- Medical expert
- Communicator
- Collaborator
- Manager
- Health advocate
- Scholar
- Professional



ACGME

- Medical knowledge
- Patient care
- Practice-based learning & improvement
- Interpersonal and communication skills
- Professionalism
- Systems-based practice



GMC

- Good clinical care
- Relationships with patients and families
- Working with colleagues
- Managing the workplace
- Social responsibility and accountability
- Professionalism

Laten we de oude situatie eens onder de loep nemen

- Trekken zijn stabiele en generieke eigenschappen
- Individuele items zijn op zichzelf betekenisloos
- Totalscores bepalen wat een toets 'meet'
- Statistiek is gebaseerd op weggooien van informatie
- Eén beste toetsmethode voor iedere trek

Trekken zijn stabiele en generiek eigenschappen

kennis

vaardigheden

probleemoplossen

attitude

TOETS

TOETS

TOETS

TOETS

TOETS

TOETS

TOETS

TOETS

toetsing: theoretische situatie

TOETS

	1	2	3	4	R
A	4	4	4	4	4
B	3.2	3.2	3.2	3.2	3.2
C	2.4	2.4	2.4	2.4	2.4
D	1.6	1.6	1.6	1.6	1.6
E	0.8	0.8	0.8	0.8	0.8

PARALLELE TOETS

	1'	2'	3'	4'	R'
A	4	4	4	4	4
B	3.2	3.2	3.2	3.2	3.2
C	2.4	2.4	2.4	2.4	2.4
D	1.6	1.6	1.6	1.6	1.6
E	0.8	0.8	0.8	0.8	0.8

toetsing: werkelijke situatie

TOETS

	1	2	3	4	R
A	Green	Yellow	Red	Green	3.2
B	Green	Orange	Dark Red	Yellow	2.6
C	Orange	Red	Green	Orange	2.6
D	Green	Yellow	Red	Orange	2.6
E	Orange	Dark Red	Red	Dark Red	1.4

PARALLELE TOETS

	1	2	3	4	R
A	Yellow	Red	Red	Green	2.6
B	Yellow	Green	Green	Orange	3.2
C	Orange	Green	Orange	Orange	2.8
D	Orange	Dark Red	Green	Red	2.2
E	Green	Red	Yellow	Dark Red	2.4

Laten we de oude situatie eens onder de loep nemen

- Trekken zijn stabiele en generieke eigenschappen
- Individuele items zijn op zichzelf betekenisloos
- Totalscores bepalen wat een toets 'meet'
- Statistiek is gebaseerd op weggooien van informatie
- Eén beste toetsmethode voor iedere trek

Individuele items zijn betekenisloos op zichzelf

Mr. Jansen is 35 jaar oud. Hij bezoekt zijn huisarts met klachten van pijn op de borst. De pijn komt en gaat en is weg te zuchten of door van houding te wisselen. Wat is op dit moment de meest waarschijnlijke oorsprong van zijn pijn?

- a de thoraxwand;
- b de longen;
- c het hart;
- d de Isokdarm.

Individuele items zijn betekenisloos op zichzelf

Dhr Jansen is 35 jaar out. Hij bezoekt zijn huisarts met klachten van pijn op de borst. De klachten komen spontaan maar zijn te beïnvloeden door te zuchten of een andere houding aan te nemen.

Wat is de meest waarschijnlijke origine van zijn klachten?

- a de borstkas;
- b de longen;
- c het hart;
- d de slokdarm.

Individuele items zijn betekenisloos op zichzelf

resuscitatie 'station'
in een
vaardigheidstoets

Individuele items zijn betekenisloos op zichzelf

communicatie
'station'
in een
vaardigheidstoets

Laten we de oude situatie eens onder de loep nemen

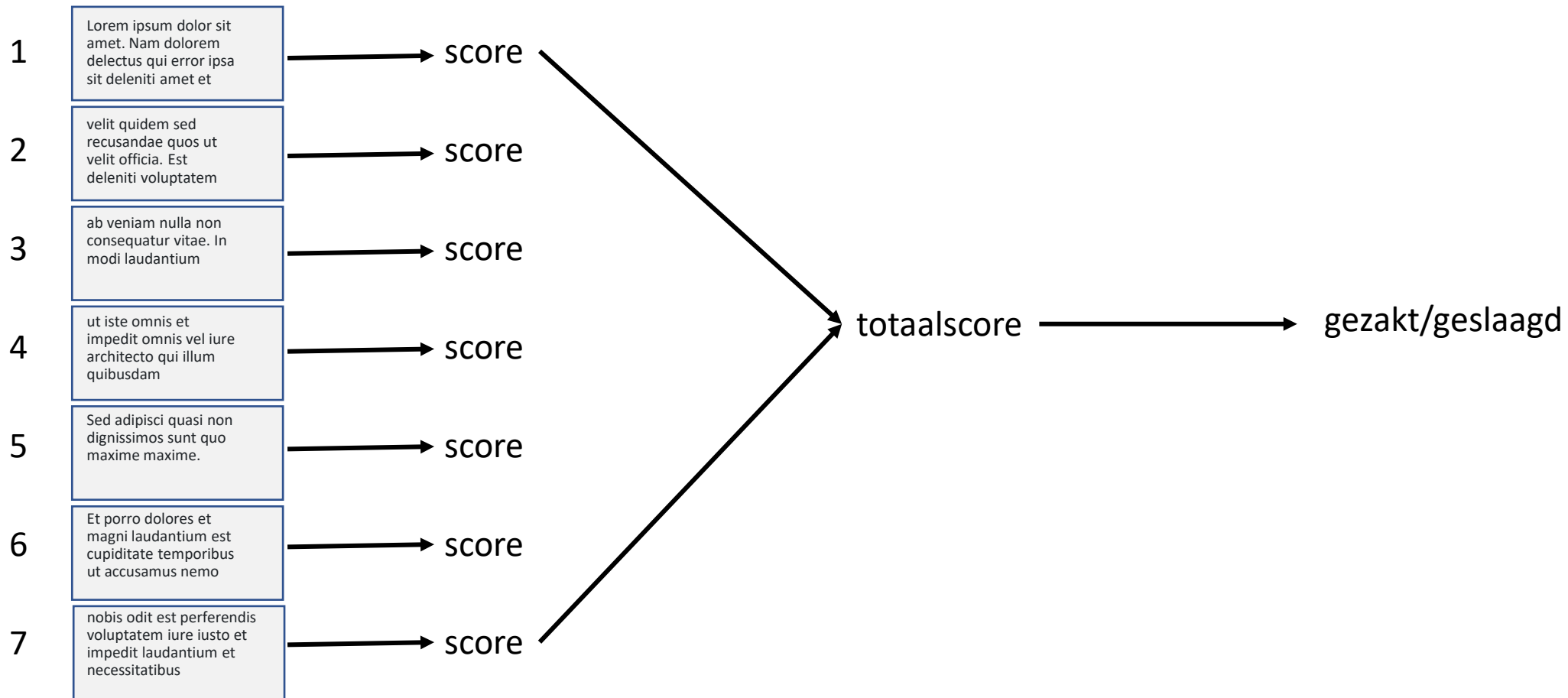
- Trekken zijn stabiele en generieke eigenschappen
- Individuele items zijn op zichzelf betekenisloos
- Totalscores bepalen wat een toets 'meet'
- Statistiek is gebaseerd op weggooien van informatie
- Eén beste toetsmethode voor iedere trek

totaalscores

Laten we de oude situatie eens onder de loep nemen

- Trekken zijn stabiele en generieke eigenschappen
- Individuele items zijn op zichzelf betekenisloos
- Totalscores bepalen wat een toets 'meet'
- Statistiek is gebaseerd op weggooien van informatie
- Eén beste toetsmethode voor iedere trek

elimination of information



Laten we de oude situatie eens onder de loep nemen

- Trekken zijn stabiele en generieke eigenschappen
- Individuele items zijn op zichzelf betekenisloos
- Totalscores bepalen wat een toets 'meet'
- Statistiek is gebaseerd op weggooien van informatie
- Eén beste toetsmethode voor iedere trek

old model of medical competence

knowledge

skills

problem solving

attitudes

TEST > TEST

TEST > TEST

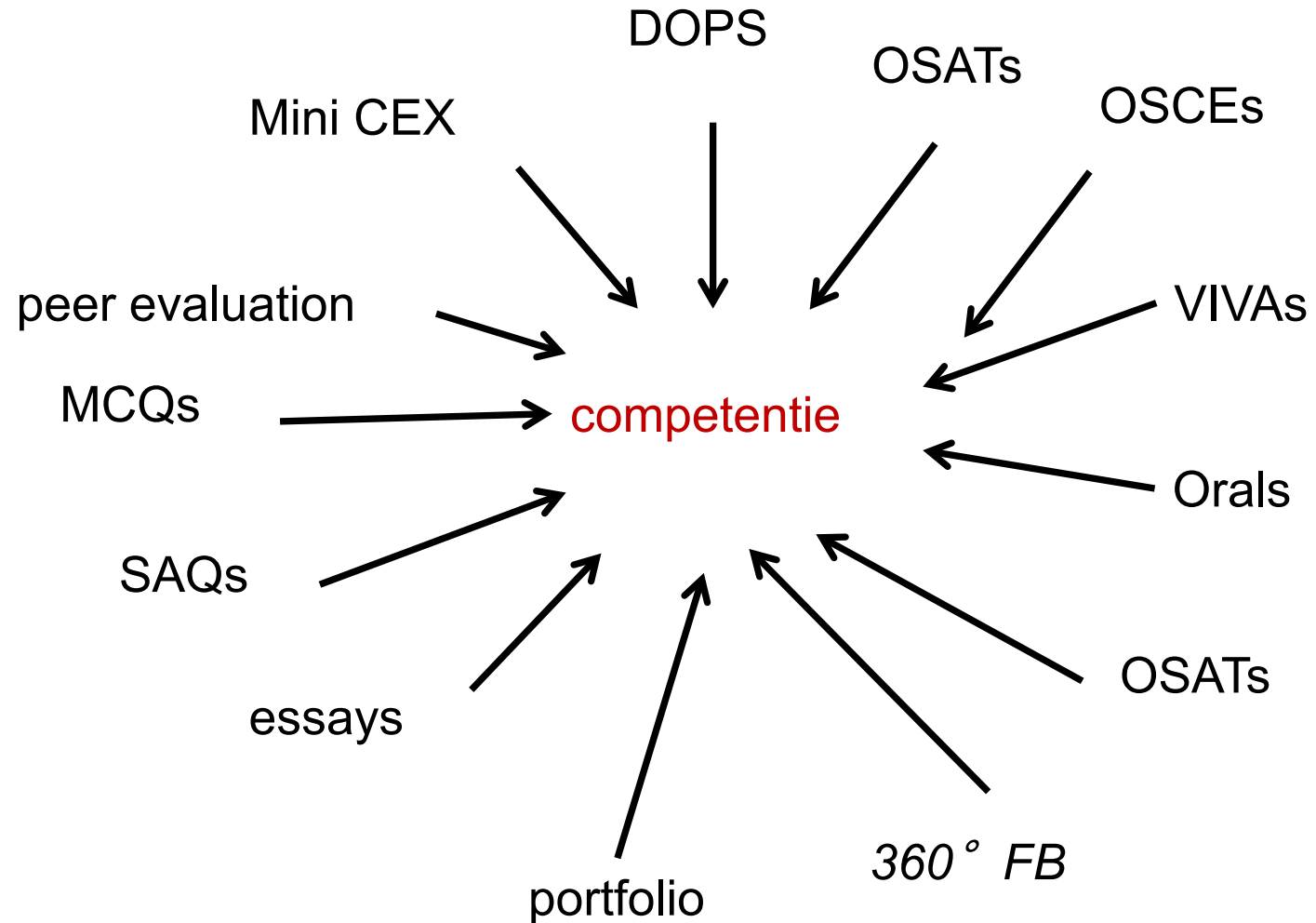
TEST > TEST

TEST > TEST

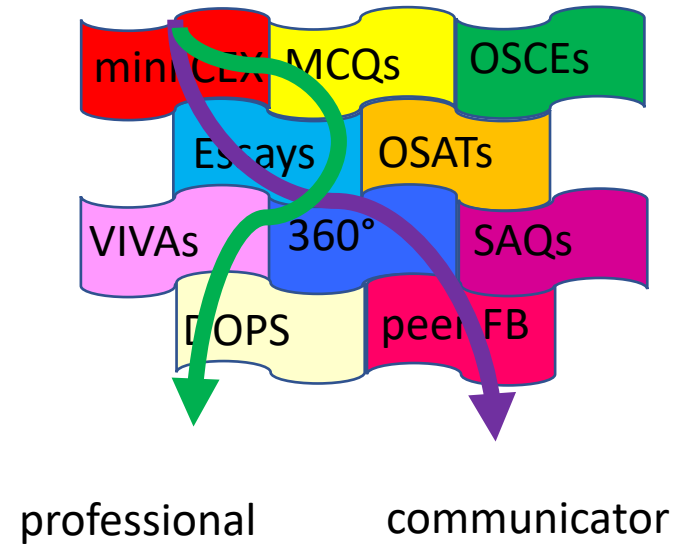
Programmatisch toetsen

- Meerdere instrumenten met betekenisvolle synthese
- Continue en longitudinale toetsing
- Proportionaliteit van beslissingen
- Variatie aan kwaliteitsdenken
- Integratief/holistisch

Meerdere instrumenten



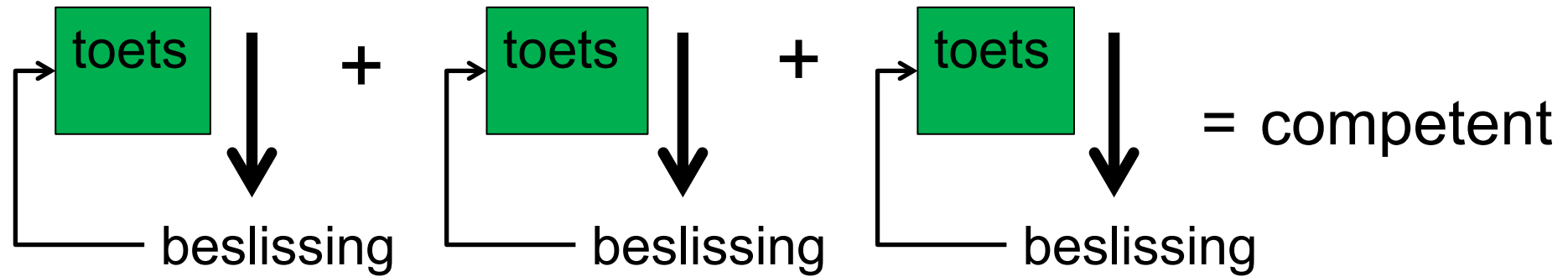
Meerdere instrumenten



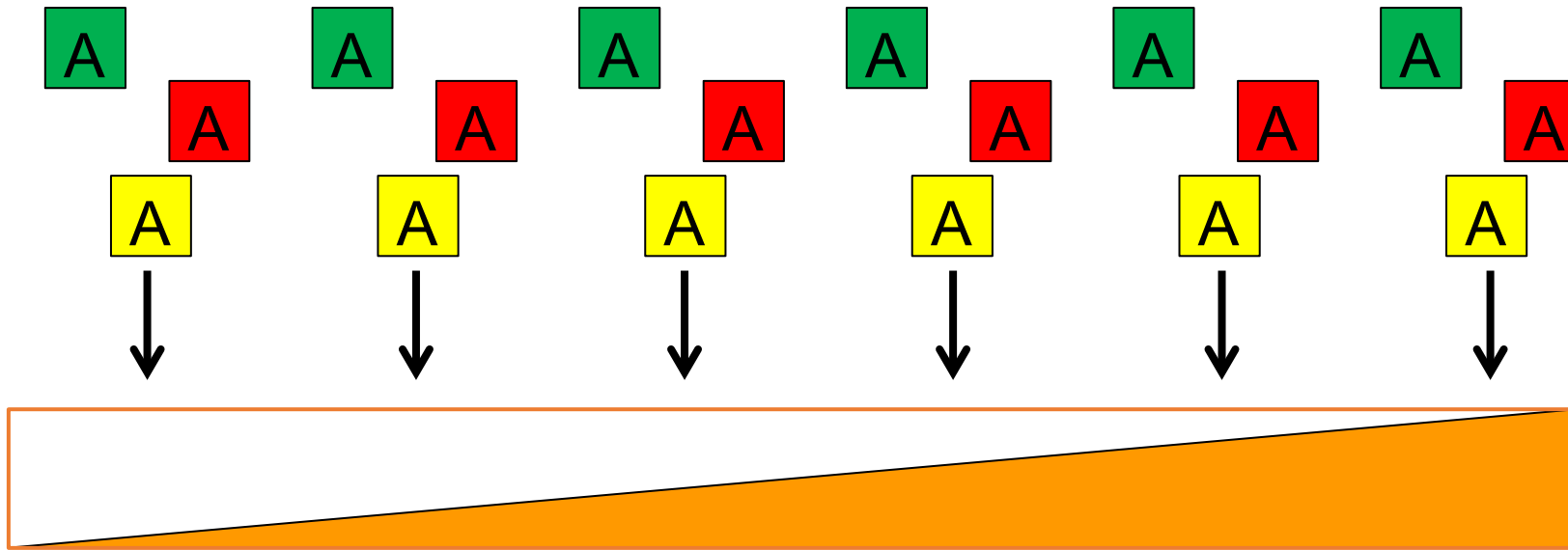
Programmatisch toetsen

- Meerdere instrumenten met betekenisvolle synthese
- Continue en longitudinale toetsing
- Proportionaliteit van beslissingen
- Variatie aan kwaliteitsdenken
- Integratief/holistisch

leder toetsmoment = beslismoment



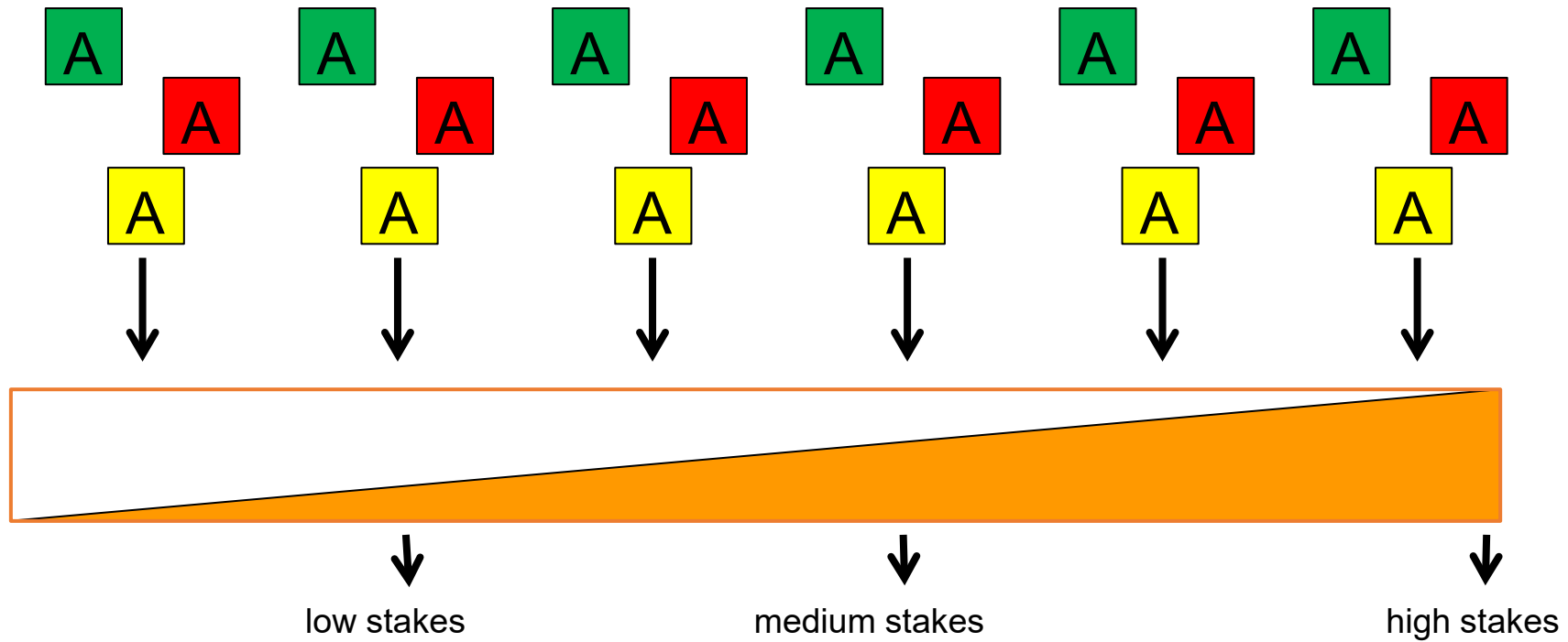
Longitudinale toetsing



Programmatisch toetsen

- Meerdere instrumenten met betekenisvolle synthese
- Continue en longitudinale toetsing
- Proportionaliteit van beslissingen
- Variatie aan kwaliteitsdenken
- Integratief/holistisch

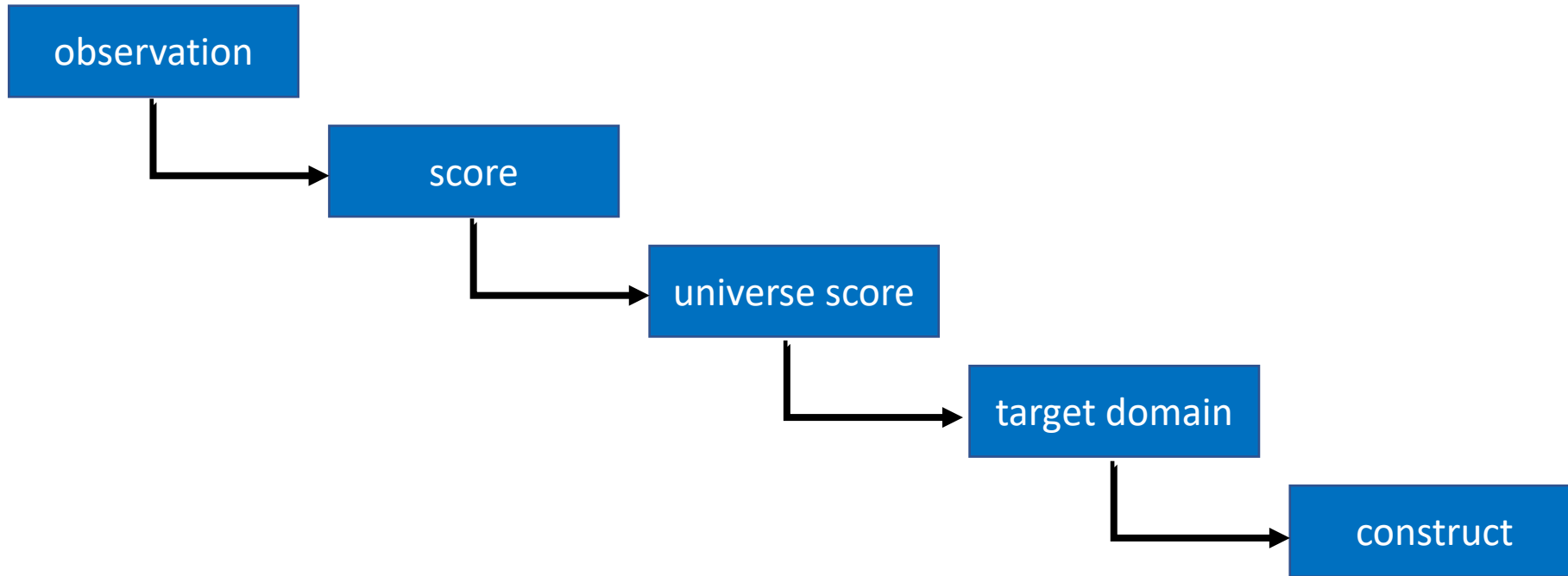
Proportionaliteit



Programmatisch toetsen

- Meerdere instrumenten met betekenisvolle synthese
- Continue en longitudinale toetsing
- Proportionaliteit van beslissingen
- Variatie aan kwaliteitsdenken
- Integratief/holistisch

validity als een keten van inferenties



kwaliteit: betrouwbaarheid

toets

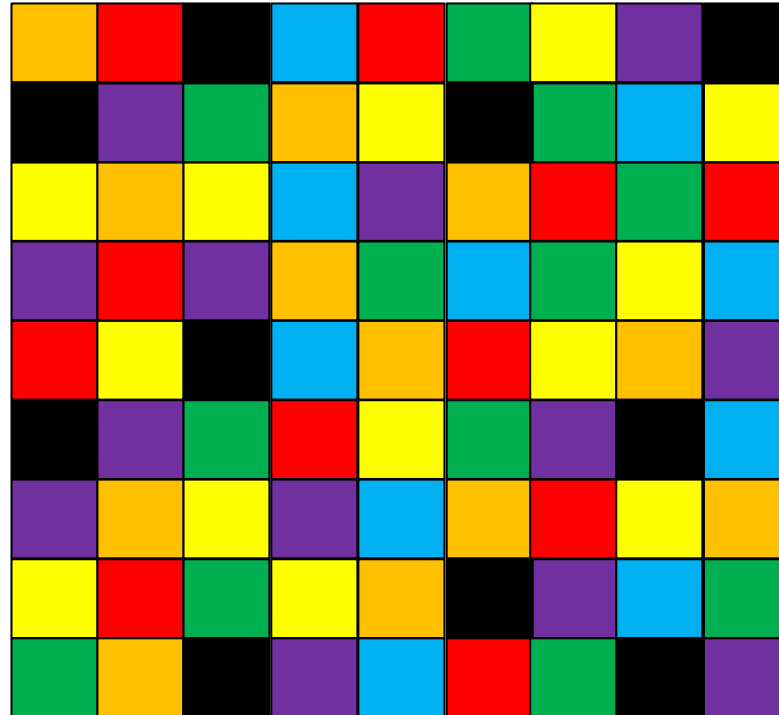
=

parallele toets



- zelfde scores
- zelfde rangorde
- zelfde beslissing

Kwaliteit: saturatie



oranje

groen

blauw

rood

geel

paars

zwart

niets nieuws

niets nieuws

Kwaliteit: fairness (rechtvaardigheid; eerlijkheid)



Overeengekomen doelstelling

Behoeften van de lerende

Theorie- en evidence-gebaseerde aannames

Procedurele rechtvaardigheid

Zinvolle verzameling en synthese

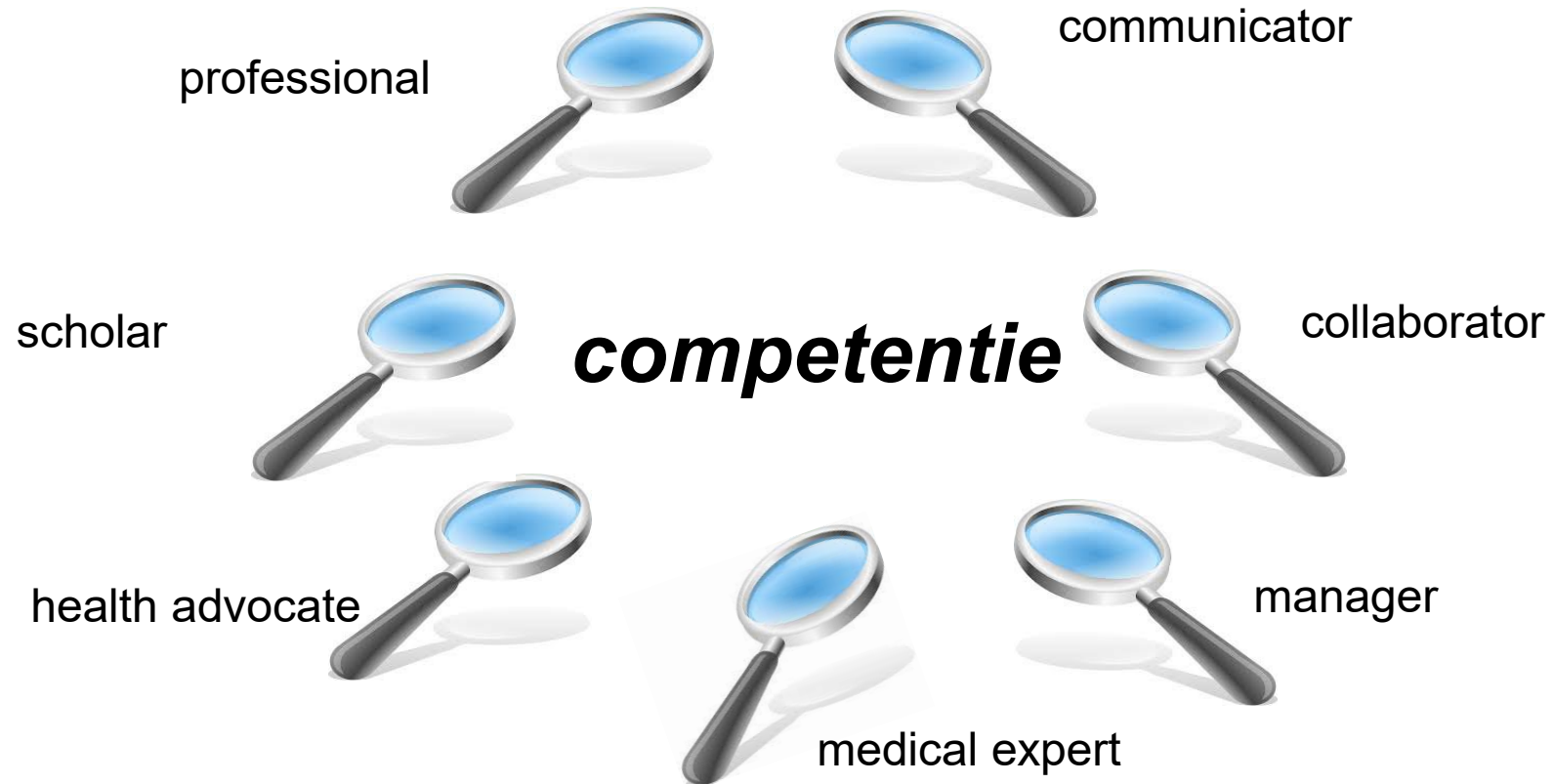
Communicatie met duidelijk narratief

Assessor flexibiliteit en expertise

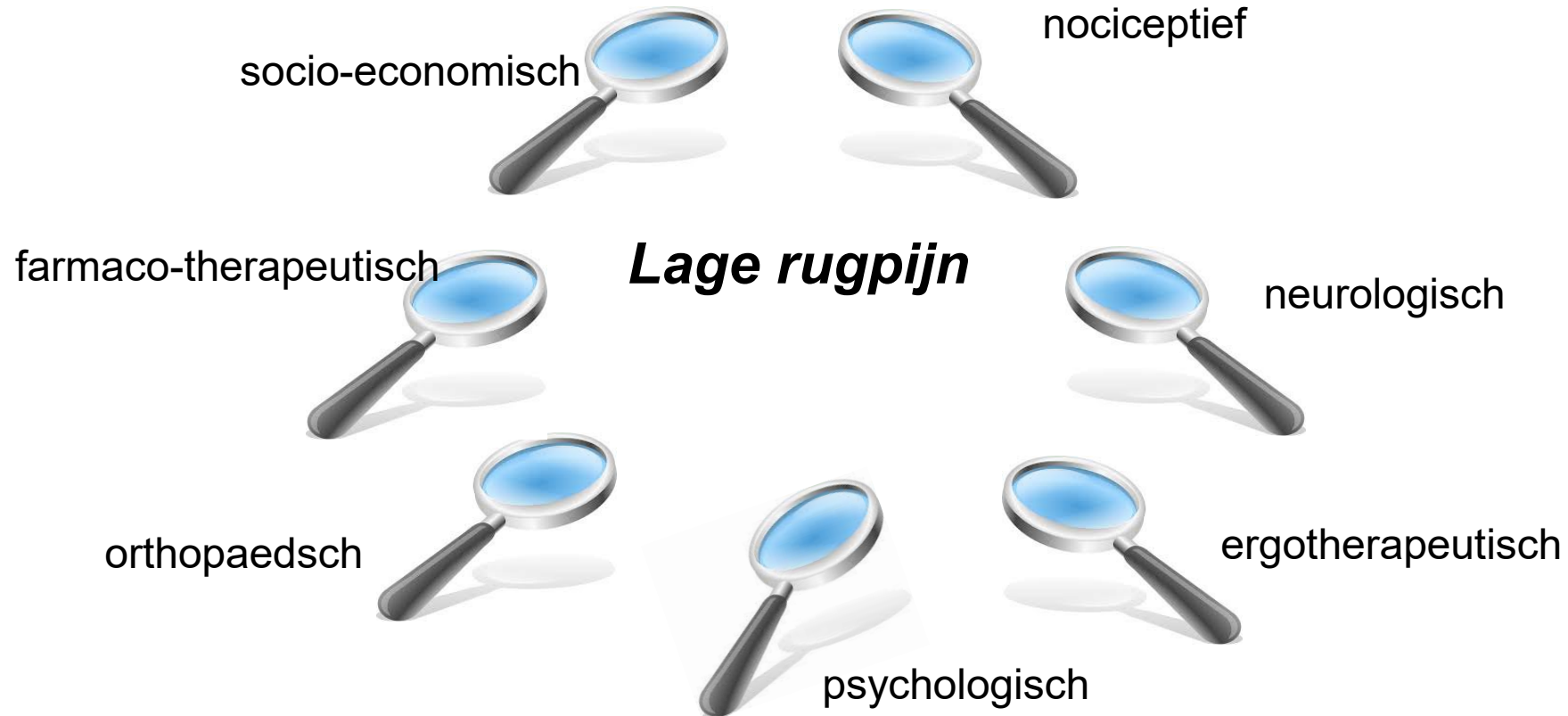
Programmatisch toetsen

- Meerdere instrumenten met betekenisvolle synthese
- Continue en longitudinale toetsing
- Proportionaliteit van beslissingen
- Variatie aan kwaliteitsdenken
- Integratief/holistisch

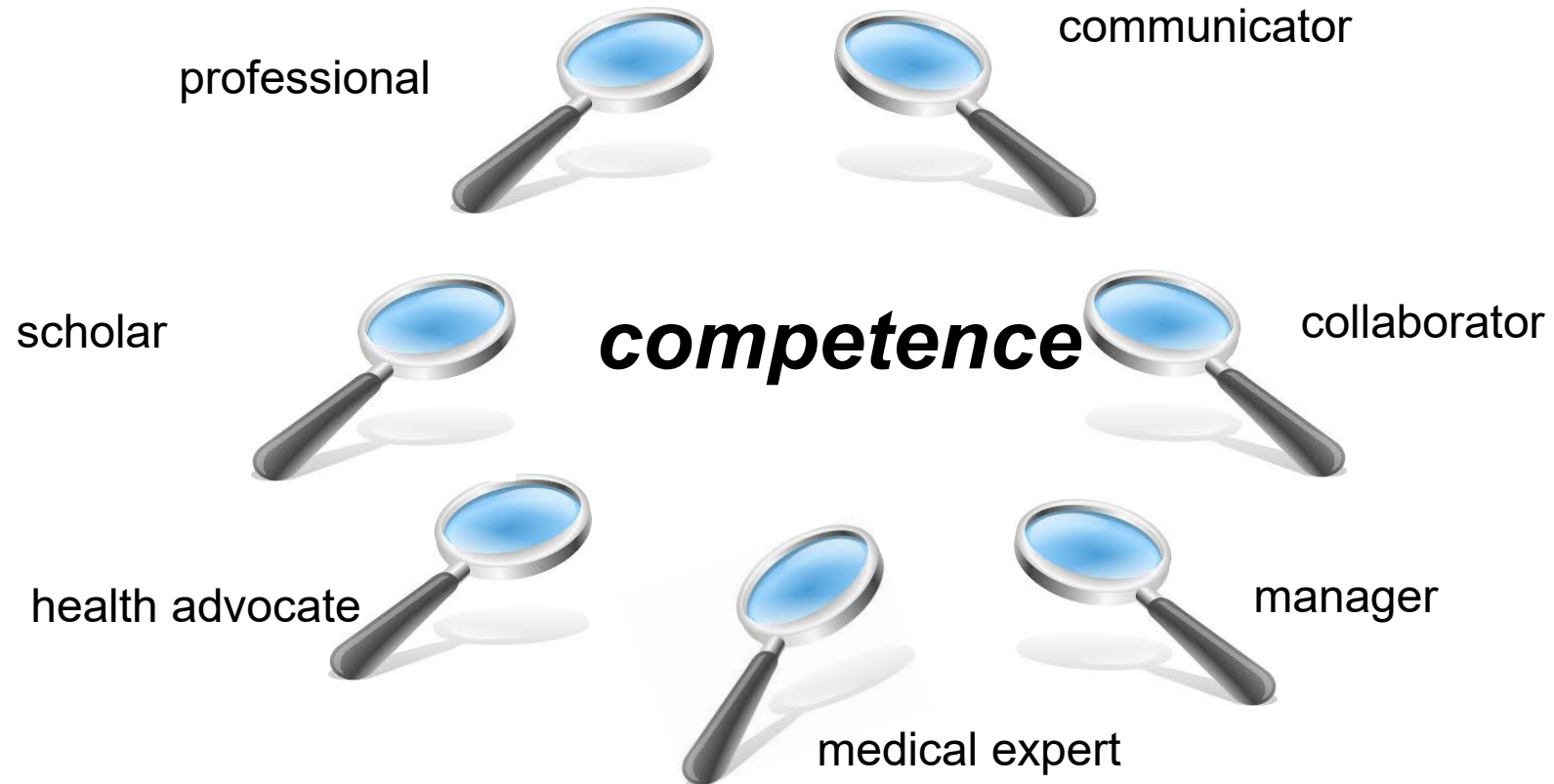
Domeinen als perspectieven



Domeinen als perspectieven



Domeinen als perspectieven

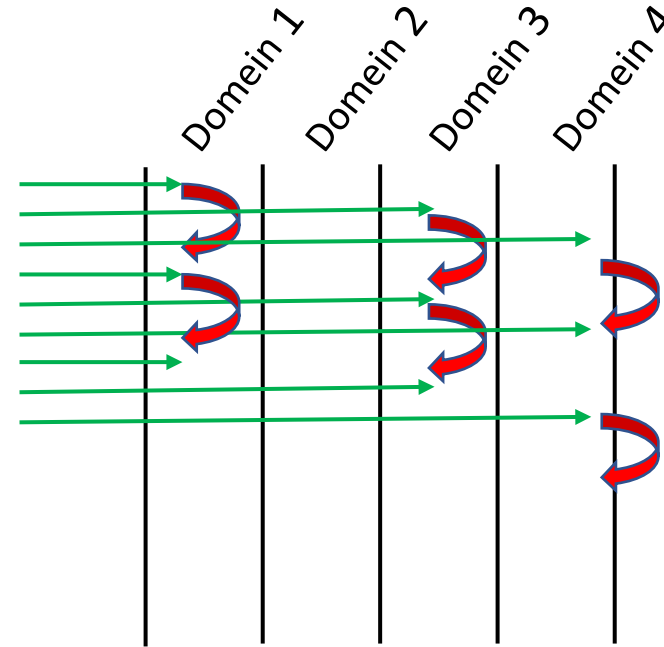
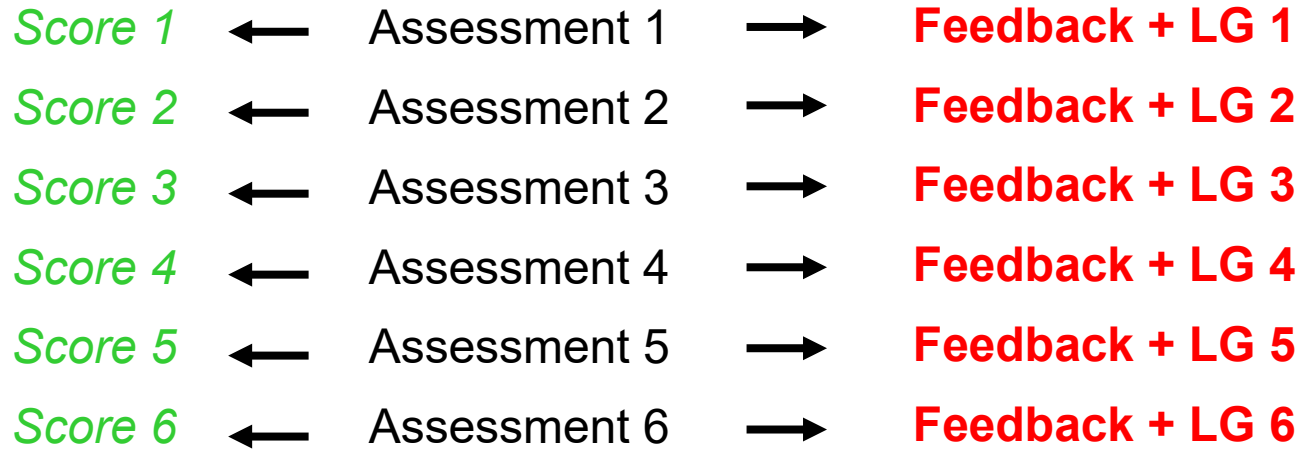


assessment for learning

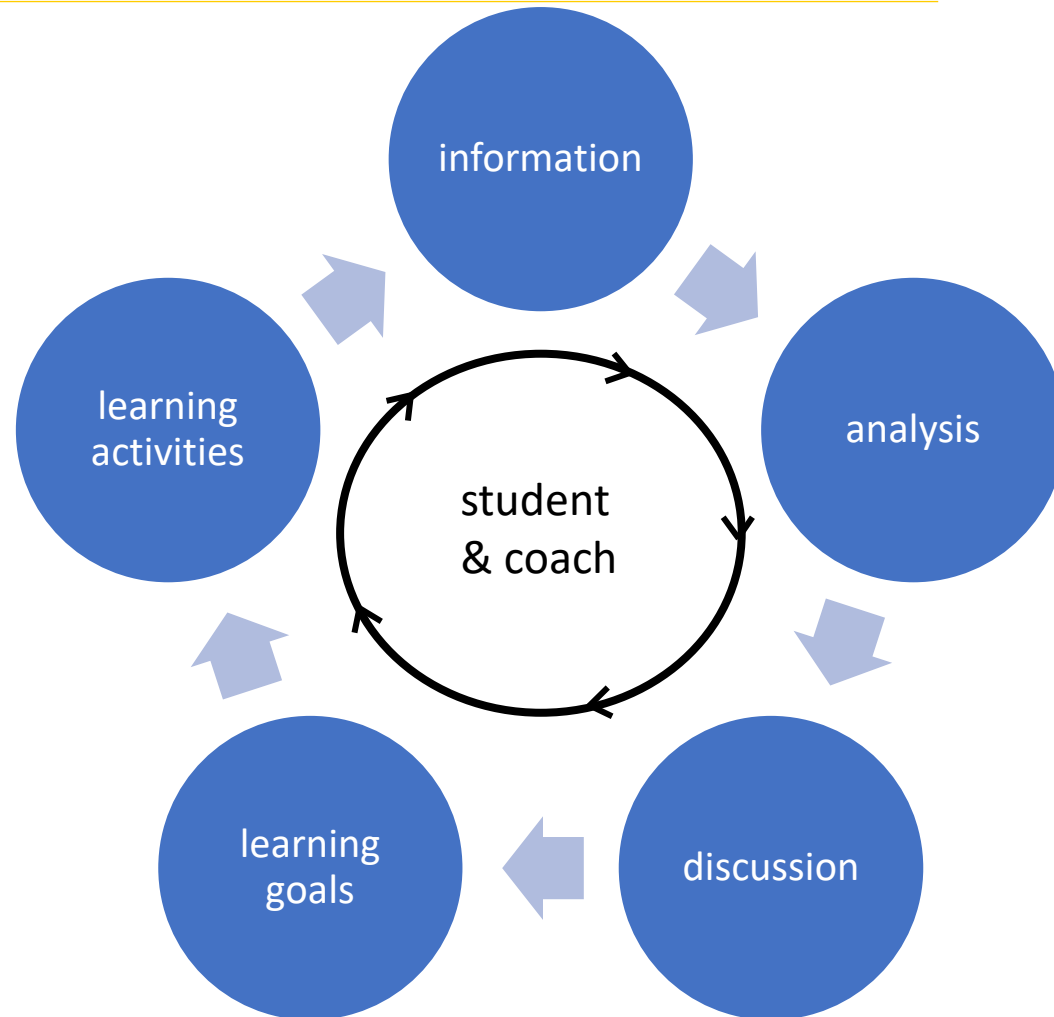


- meaningfulness
- feedback
- agency
- responsibility

Assessment for learning



cycles



*Thank
you*



background reading

1. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological bulletin* 1955;52(4):281 - 302.
2. Ebel RL. The practical validation of tests of ability. *Educational Measurement: Issues and Practice* 1983;2(2):7 - 10.
3. Cronbach LJ. What price simplicity? *Educational Measurement: Issues and Practice* 1983;2(2):11-12.
4. Kane MT. Validation. In: Brennan RL, ed. *Educational Measurement*. Westport: ACE/Praeger 2006:17 - 64.
5. Norman GR, Smith EKM, Powles AC, et al. Factors underlying performance on written tests of knowledge. *Medical Education* 1987;21:297-304.
6. Ward WC. A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Applied Psychological Measurement* 1982;6(1):1-11.
7. Van der Vleuten CPM, Van Luyk SJ, Beckers HJM. A written test as an alternative to performance testing. *Medical Education* 1988;22:97-107.
8. Norman G, Swanson D, Case S. Conceptual and methodology issues in studies comparing assessment formats, issues in comparing item formats. *Teaching and Learning in Medicine* 1996;8(4):208-16.
9. Norman GR, Van der Vleuten CP, De Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Medical Education* 1991;25(2):119-26.
10. Boud D. Assessment and the promotion of academic values. *Studies in Higher Education* 1990;15(1):101-11.
11. Govaerts MJB, Van der Vleuten CPM, Schuwirth LWT, et al. Broadening Perspectives on Clinical Performance Assessment: Rethinking the Nature of In-training Assessment. *Advances in health sciences education* 2007;12(2):239-60.
12. Canmeds. <http://www.royalcollege.ca/portal/page/portal/rc/canmeds> Ottawa2005 [accessed 26 August 2018].
13. ACGME. <http://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview> Chicago2007 [accessed August 2018].
14. Albanese MA, Mejicano G, Mullan P, et al. Defining characteristics of educational competencies. *Medical Education* 2008;42(3):248-55.

background reading

15. Van der Vleuten CPM. The assessment of Professional Competence: Developments, Research and Practical Implications. *Advances in Health Science Education* 1996;1(1):41-67.
16. Kane M. Current concerns in validity theory. *Journal of Educational Measurement* 2001;38(4):319-42.
17. Schuwirth LWT, Van der Vleuten CPM. A plea for new psychometrical models in educational assessment. *Medical Education* 2006;40(4):296-300.
18. Driessen E, Van der Vleuten CPM, Schuwirth LWT, et al. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 2005;39(2):214-20.
19. Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Medical Education* 2005;39(3):309-17.
20. Dannefer E, Henson L. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007;82(5):493-502.
21. Van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Medical Teacher* 2012;34:205-14. doi: 10.3109/0142159X.2012.652239
22. Schuwirth LWT, Van der Vleuten CPM, Durning SJ. What programmatic assessment for learning in medical education can learn from healthcare. *Perspectives on Medical Education* 2017:1-5. doi: 10.1007/s40037-017-0345-1
23. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. *Medical Education* 2012;46(1):38-48. doi: 10.1111/j.1365-2923.2011.04098.x
24. Schuwirth LWT, Van der Vleuten CPM. Assessing Competence: Extending the Approaches to Reliability. In: Hodges BD, Lingard L, eds. *The Question of Competence*. Ithaca New York US: Cornell University Press 2012.
25. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher* 2011;33(6):478-85.

background reading

26. Delandshere G, Petrosky AR. Assessment of Complex Performances: Limitations of Key Measurement Assumptions. *Educational Researcher* 1998;27(2):14-24.
27. Shute V. Focus on formative feedback. *Review of educational research* 2008;78(n):153-89.
28. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Medical Education* 2007;41:1124-30. doi: 10.1111/j.1365-2923.2007.02946.x
29. Cook DA, Kuper A, Hatala R, et al. When assessment data are words: validity evidence for qualitative educational assessments. *Academic Medicine* 2016;91(10):1359-69.
30. Popham WJ. Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice* 2009;48:4-11.
31. Watling C, Driessen E, Van der Vleuten CPM, et al. Beyond individualism: professional culture and its influence on feedback. *Medical Education* 2013;47(6):585-94.
32. Harrison CJ, Könings KD, Schuwirth L, et al. Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in health sciences education* 2015;20(1):229–45. doi: 10.1007/s10459-014-9524-6
33. Cilliers FJ, Schuwirth LWT, Adendorff HJ, et al. The mechanisms of impact of summative assessment on medical students' learning. *Advances in health sciences education* 2010;15:695-715. doi: 10.1007/s10459-010-9232-9
34. Valentine N, Durning S, Shanahan EM, et al. Fairness in human judgement in assessment: A hermeneutic literature review and conceptual framework. *Advances in Health Sciences Education* 2021;26(2):713-38.
35. Valentine N, Shanahan EM, Durning SJ, et al. Making it fair: Learners' and assessors' perspectives of the attributes of fair judgement. *Medical Education* 2021;55(9):1056-66. doi: <https://doi.org/10.1111/medu.14574>
36. Valentine N, Durning SJ, Shanahan EM, et al. The pursuit of fairness in assessment: Looking beyond the objective. *Medical Teacher* 2022:1-7.