

Wirksam Lehren – Wirksam Prüfen

Eine kurze Reise im Spiegel der Lehr-
/Lernforschung

Dr. med. M. Hofmann, MME (Bern)

Prodekan für Lehre der Fakultät für Gesundheit
Universität Witten/Herdecke

Überblick & Ziele



- Von der gelebten Praxis zu den Forschungsergebnissen...
- Von den Forschungsergebnissen zu einem neue Prüfungsansatz...
- Vom neuen Prüfungsansatz zu einer veränderten Praxis...



Wie prüfen wir? Die Werkzeugkiste...

MCQ, SAQ, MEQ, Key-Feature, Progress Test, Long Case, Short Case, OSCE, OSPE, SP-based test, video-assessment, Mini-CEX, DOPS, MMI's, self-assessment, peer-assessment, portfolio, etc.



Bei testtheoretischen Qualitätskriterien & Instrumenten? Oft erhebliche Kompromisse...

Reliabilität & Prüfungsdauer...

Testing Time in Hours	MCQ ¹	Case-Based Short Essay ²	PMP ¹	Oral Exam ³	Long Case ⁴	OSCE ⁵	Mini CEX ⁶	Practice Video Assessment ⁷	In-cognito SPs ⁸
1	0.62	0.68	0.36	0.50	0.60	0.54	0.73	0.62	0.61
2	0.76	0.73	0.53	0.69	0.75	0.69	0.84	0.76	0.76
4	0.93	0.84	0.69	0.82	0.86	0.82	0.92	0.93	0.92
8	0.93	0.82	0.82	0.90	0.90	0.90	0.96	0.93	0.93

¹Norcini et al., 1985

²Stalenhoef-Halling et al., 1990

³Swanson, 1987

⁴Wass et al., 2001

⁵Van der Vleuten, 1988

⁶Norcini et al., 1999

⁷Ram et al., 1999

⁸Gorter, 2002

Was lernen wir daraus?

- Zufriedenstellende Reliabilität wird nur über großen Stichprobenumfang erreicht.
- Keine Methode ist einer anderen grundsätzlich überlegen.
- Objektivität ist nicht gleich Reliabilität!
- Viele subjektiven Einschätzungen bleiben dennoch reproduzierbar.

Was prüfen wir? Validität...



Can Meds

- Medical Expert
- Communicator
- Collaborator
- Manager
- Health Advocate
- Scholar
- Professional



ACGME

- Medical knowledge
- Patient care
- Practice-based learning
- Interpersonal & communicative skills
- Professionalism
- System-based practice



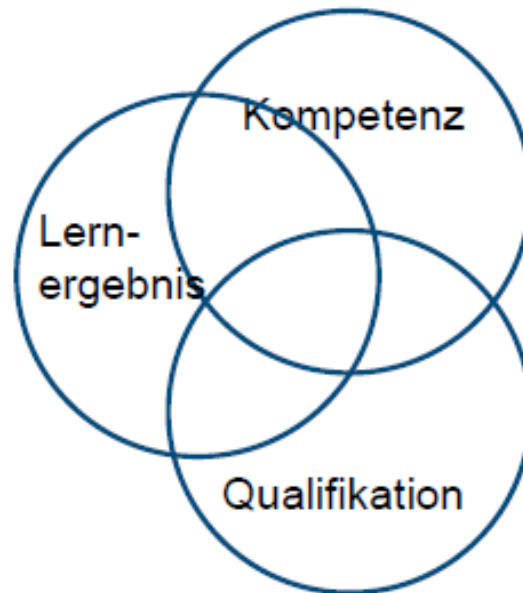
GMC

- Good clinical care
- Relationship with patients & families
- Working with colleagues
- Managing the workplace
- Social responsibility & accountability

Was prüfen wir? Validität...

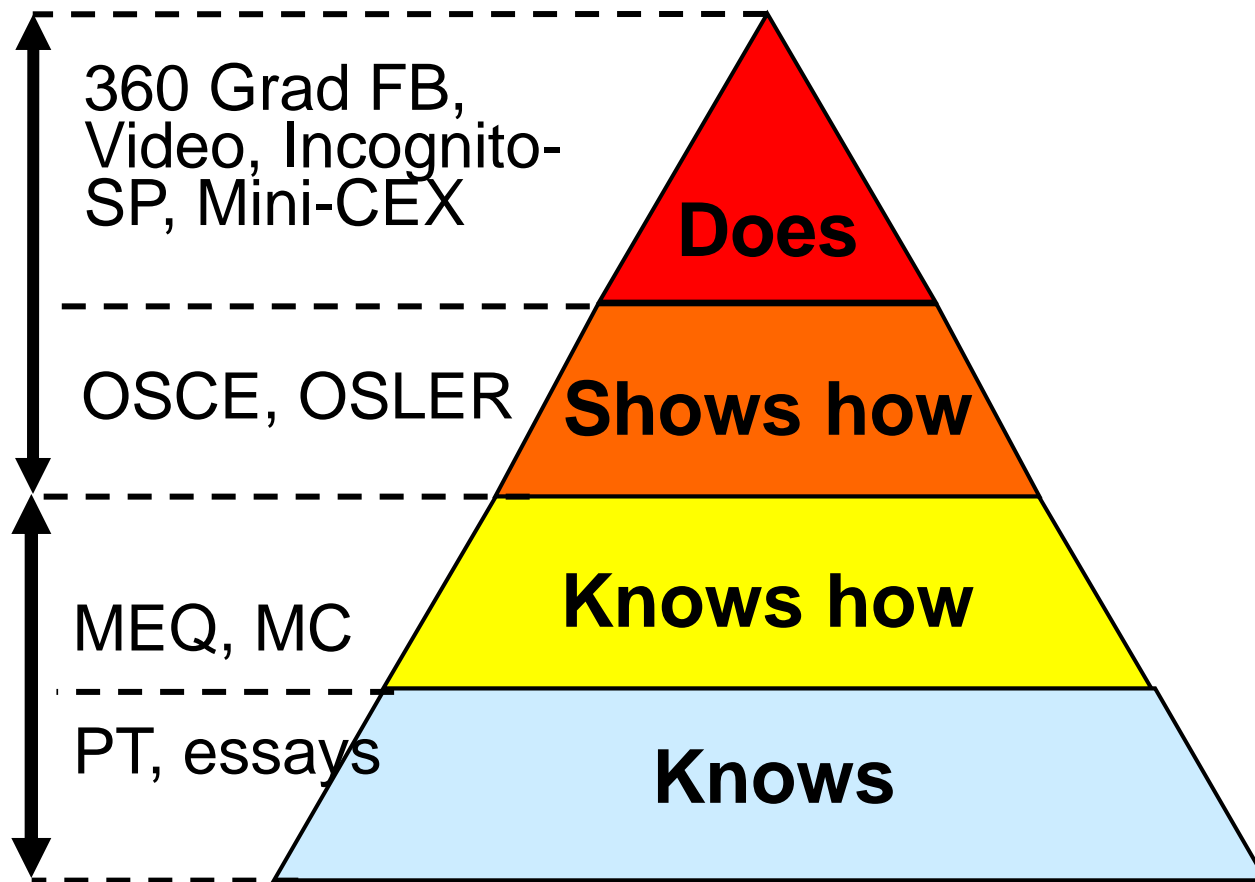
“Competence is the (individual) ability to make deliberate choices from a repertoire of behaviors to handle situations and tasks in specific contexts of professional practice, by using and integrating knowledge, skills, judgment, attitudes and personal values, in accordance with professional role and responsibilities.”

Lernziel = Lernergebnis = Kompetenz?

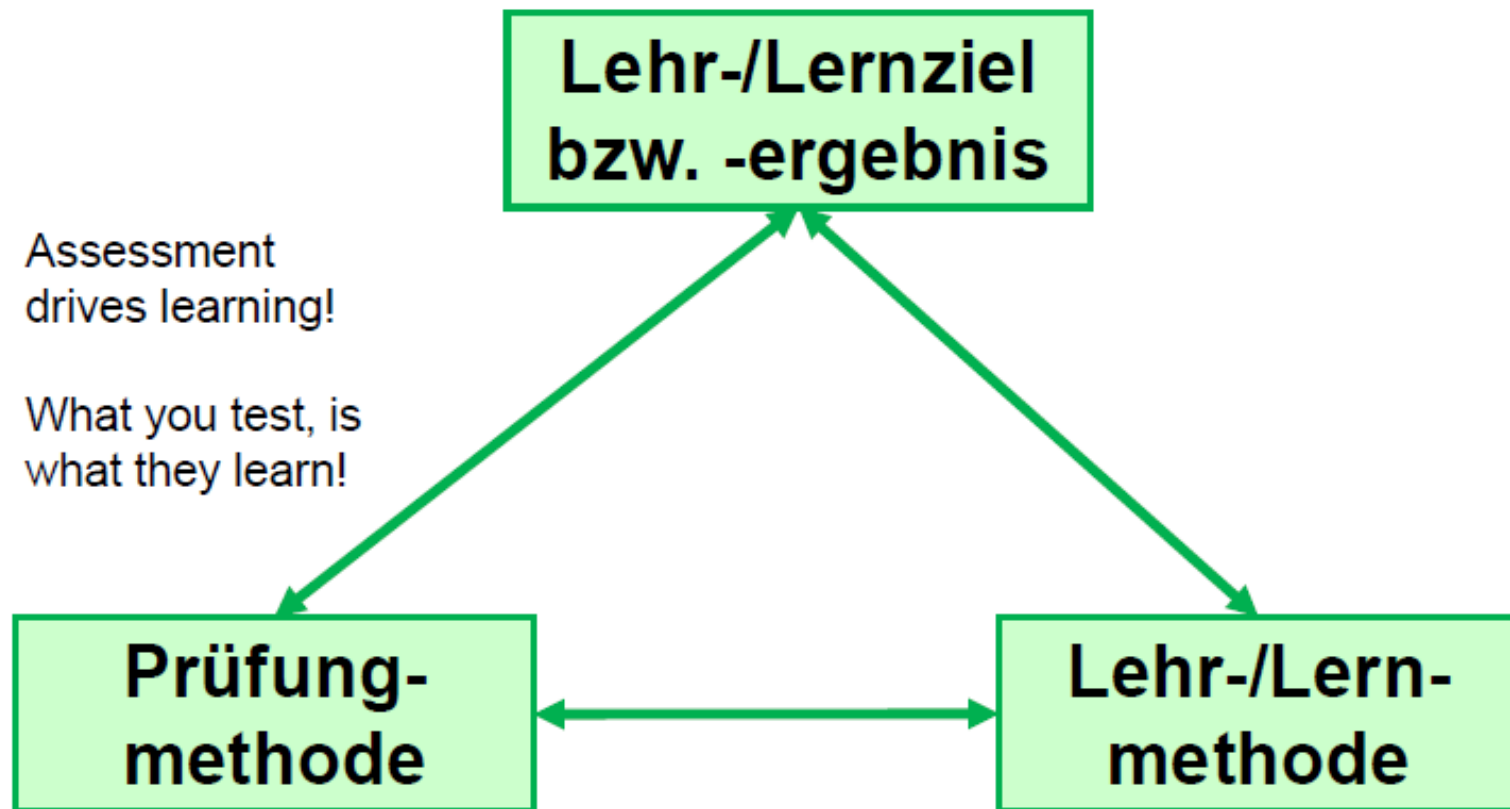


Lernergebnisse beschreiben nur den sichtbaren Teil der Kompetenz

Was prüfen wir? Validität...



„Constructive Alignment“



Tang C und Biggs J (2011): Teaching for Quality Learning at University (4th Ed.) Open University Press/ McGraw Hill, Buckingham

Validität: wenig genutztes Potential

- Self assessment
- Peer assessment
- Multisource Feedback
- Logbücher
- Critical incident reports
- Reflective diary
- **Portfolio assessment**

Was lernen wir daraus?

- Es gibt keine Allzweckwaffe – es geht um Constructive Alignment!
- Wir müssen in den Begriffen sehr genau sein...
- Das Ziel einer *wissenschaftlich und praktisch eigenverantwortlichen und selbstständigen Arztpersönlichkeit* (§1 ÄAppO) ist ein breites Konstrukt!
- Bei standardisierten Prüfungen ist die Validität von Prüfungsentwicklung und Qualitätskontrolle abhängig.
- Bei nicht-standardisierten Prüfungen ist die Validität von den beteiligten Personen abhängig.

Auswirkungen von Prüfungen

„Assesement drives learning!“

- Wheatley, Lucy; McInch, Alex; Fleming, Scott; and Lord, Rhiannon (2015) "Feeding back to feed forward: Formative assessment as a platform for effective learning," *Kentucky Journal of Higher Education Policy and Practice*: Vol. 3: Iss. 2, Article 2
- Haak et al (2011) Increased Structure and Active Learning reduce the Achievement Gap in Introductory Biology. *Science* 332, 1213
- Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach*. 2011;33(11):919–927.
- Spring et al (2011) Impact of pass/fail grading on medical students' well-being and academic outcomes, *Medical Education* 2011: 45: 867–877
- Larsen et al (2008) Test enhanced learning in medical education. *Medical Education*, 42, 959-966
- Karpicke JD et al (2008) The Critical Importance of Retrieval for Learning. *Science* 319, 96
- Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. The use of observational diaries in in-training evaluation: student perceptions. *Adv Health Sci Educ Theory Pract*. 2005;10(3):171–188.
- Delandshere G, Petrosky AR. Assessment of complex performances: limitations of key measurement assumptions. *Educ Res* 1998;27 (2):14–24.

Auswirkungen von Prüfungen

Aufgabe

Oberflächenansatz (surface-level approach):

- Auswendiglernen spezifischer Fakten und isolierter Informationsteile
- Reproduktions-Orientierung

Strategischer Ansatz

Tiefenansatz (deep-level approach):

- Suche nach Verbindungen zwischen verschiedenen Aspekten des Textes, zwischen Text und Umwelt, Vorwissen, etc.
- Bedeutungs-Orientierung

Einfluss der Lehre auf gutes Lernen

förderlich

- berufliche Relevanz erkennbar
- Sinnvolle Lernziele
- Selbst bestimmtes Lernen, problemorientiertes Lernen
- Übungsmöglichkeiten
- Feedback/ formative Prüfungen

hinderlich

- Viel Stoff, Druck, wenig Zeit
- Keine Lernziele
- Fragliche Relevanz
- fehlendes Feedback durch Peers
- „Kochrezept-Lehre“, Checklisten...
- summative Prüfung, **NOTEN**

Übergeordnete Faktoren:

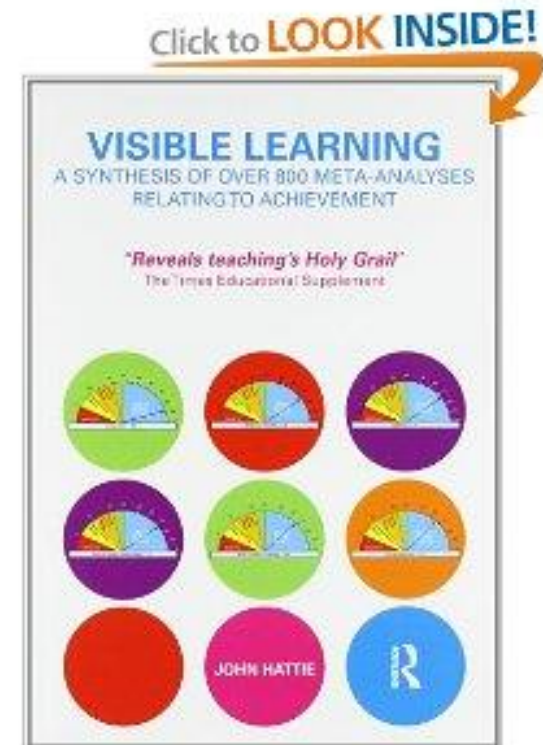
- Angst (führt zur „Regression“ in ungünstige Lernstrategien)
- Wahrnehmung der Lernenden zählt!

Mattick & Knight 2007

Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance, r. Butler, Br. J. educ. Psychol., 58, 1-14, 1988

John Hattie: What makes a difference?

- Innovation (reflection & willingness to change perspectives)
- **Feedback!!!** (reinforcing, correcting)
- Goal setting (appropriate, specific, challenging)
- Teachers as Coaches





Von den Forschungsergebnissen zu einem neuen Prüfungsansatz

- Keine Prüfungen ohne Feed-Up, Feedback & Feed-Forward (J. Hattie).
- Narratives Feedback hat einen deutlich stärkeren Einfluss auf komplexe Fähigkeiten & Fertigkeiten als Noten.
- Wir brauchen multiple Methoden, große Stichproben und longitudinale formative Prüfungen, die bedeutungsorientiertes Tiefenlernen ermöglichen.

Wheatley, Lucy; McInch, Alex; Fleming, Scott; and Lord, Rhiannon (2015) "Feeding back to feed forward: Formative assessment as a platform for effective learning," *Kentucky Journal of Higher Education Policy and Practice*: Vol. 3: Iss. 2, Article 2.

Von den Forschungsergebnissen zu einem neuen Prüfungskonzept

TWELVE TIPS

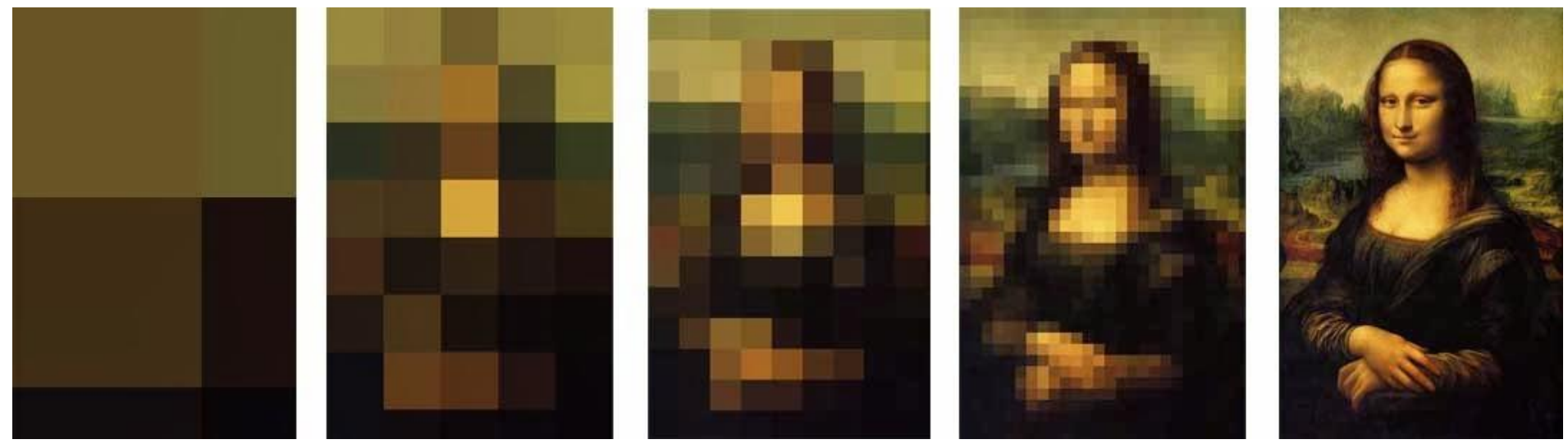
Twelve Tips for programmatic assessment

C.P.M. VAN DER VLEUTEN¹, L.W.T. SCHUWIRTH², E.W. DRIESSEN¹, M.J.B. GOVAERTS¹ & S. HEENEMAN¹

¹Maastricht University, Maastricht, The Netherlands, ²Flinders University, Adelaide, Australia

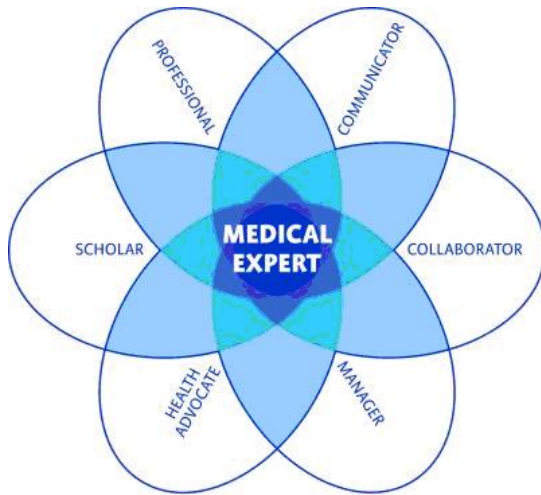
*Programmatic assessment is an integral approach to the design of an assessment (...). Individual methods of assessment, purposefully chosen for their alignment (...) are seen as individual data points. The information value (...) is maximised by giving feedback to the learner. **There is a decoupling of assessment moment and decision moment.** Intermediate and high-stakes decisions are based on multiple data points (...). Self-regulation of learning, through analysis of the assessment information and the attainment of the ensuing learning goals, is scaffolded by a mentoring system.*

Von den Forschungsergebnissen zu einem neuen Prüfungskonzept

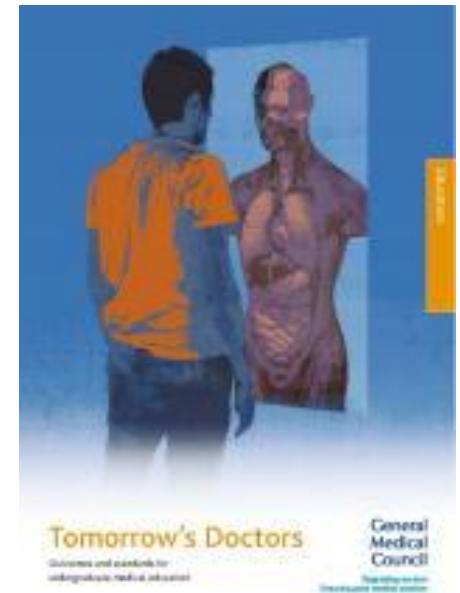
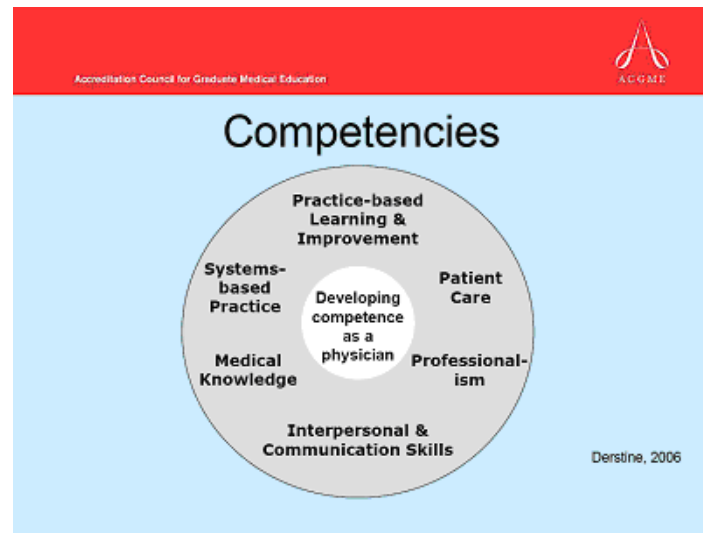




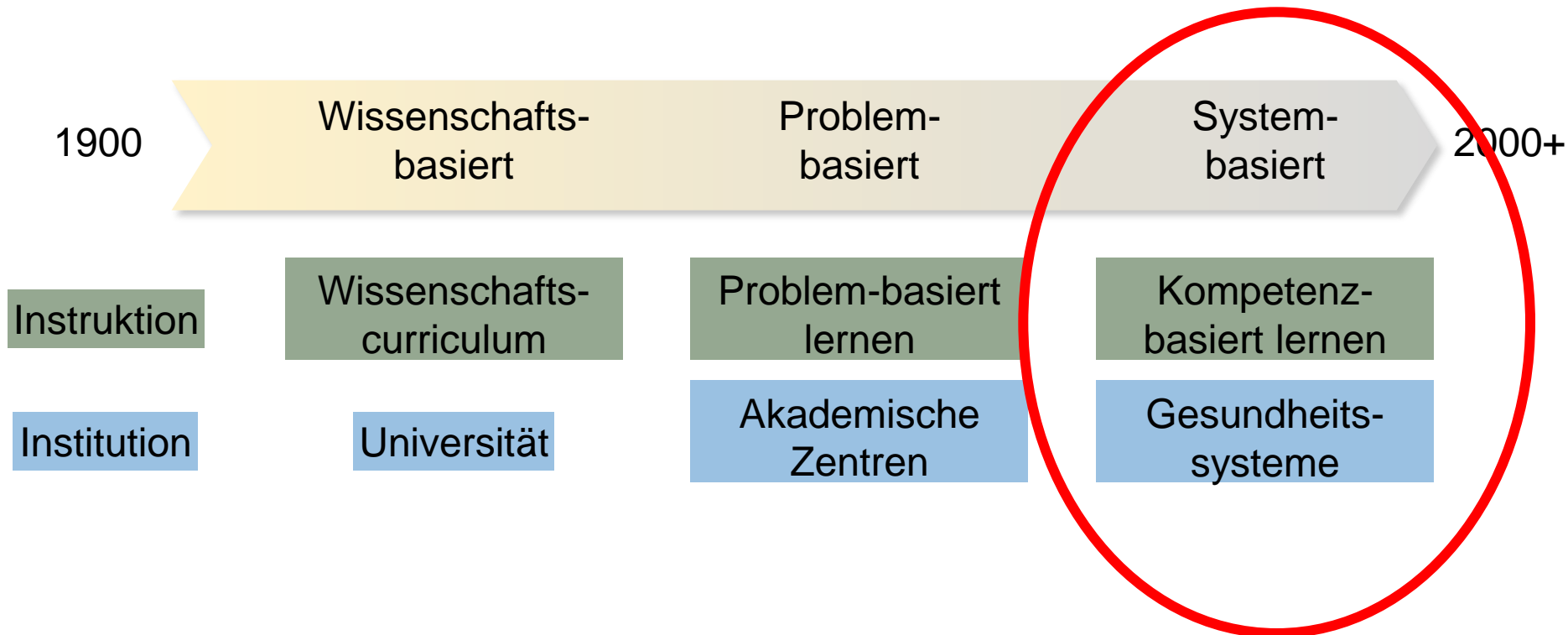
Vom neuen Prüfungskonzept zu einer verwandelten Praxis



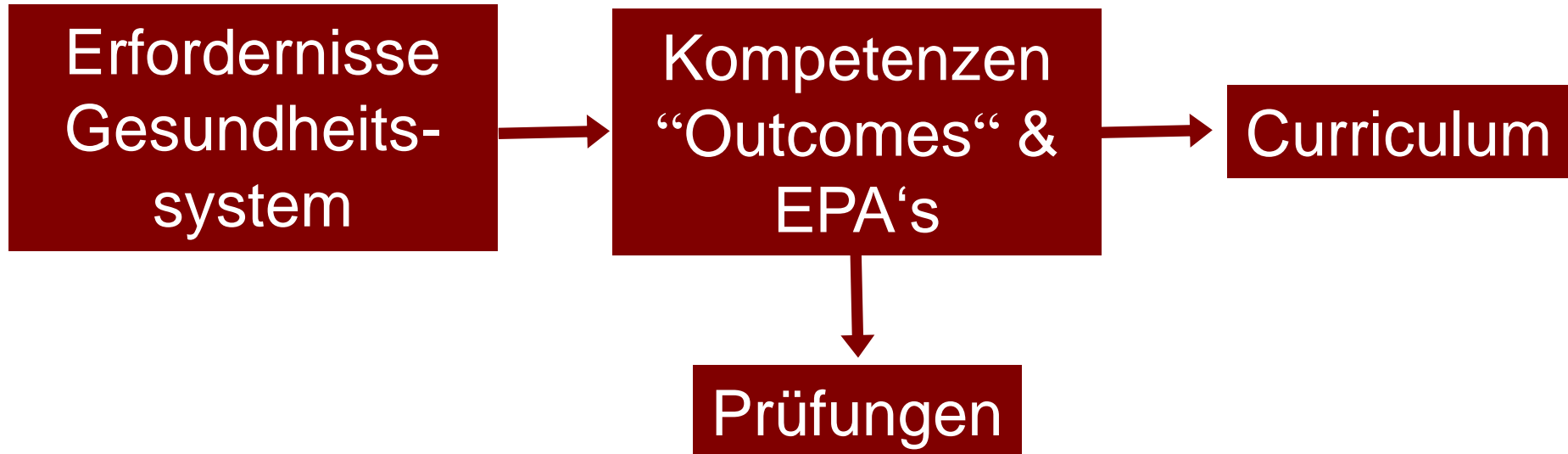
THE
CANMEDS
ROLES FRAMEWORK



Medizinische Ausbildung im Wandel der Zeit



Perspektivwechsel & „constructive alignment“



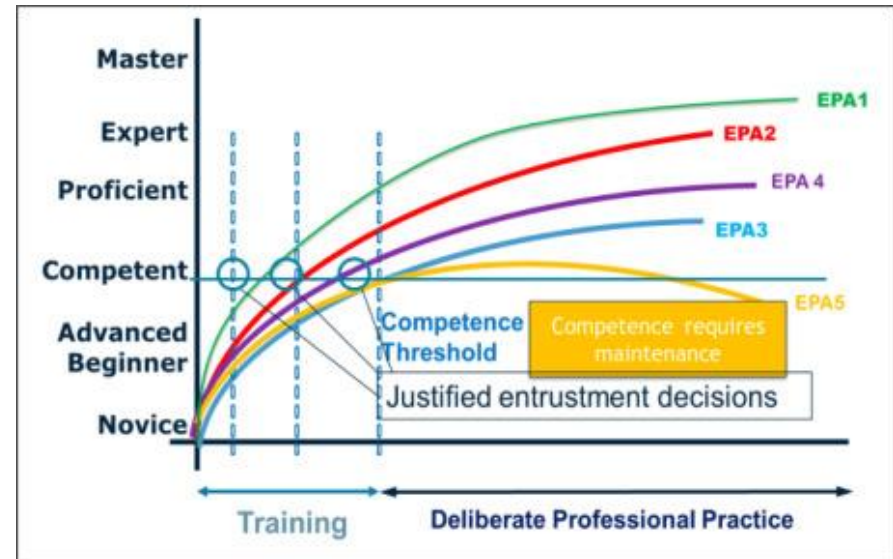
CBME, EPA's & Programmatic assessment

COMBINING THE COMPETENCY-FRAMEWORK WITH ENTRUSTABLE PROFESSIONAL ACTIVITIES

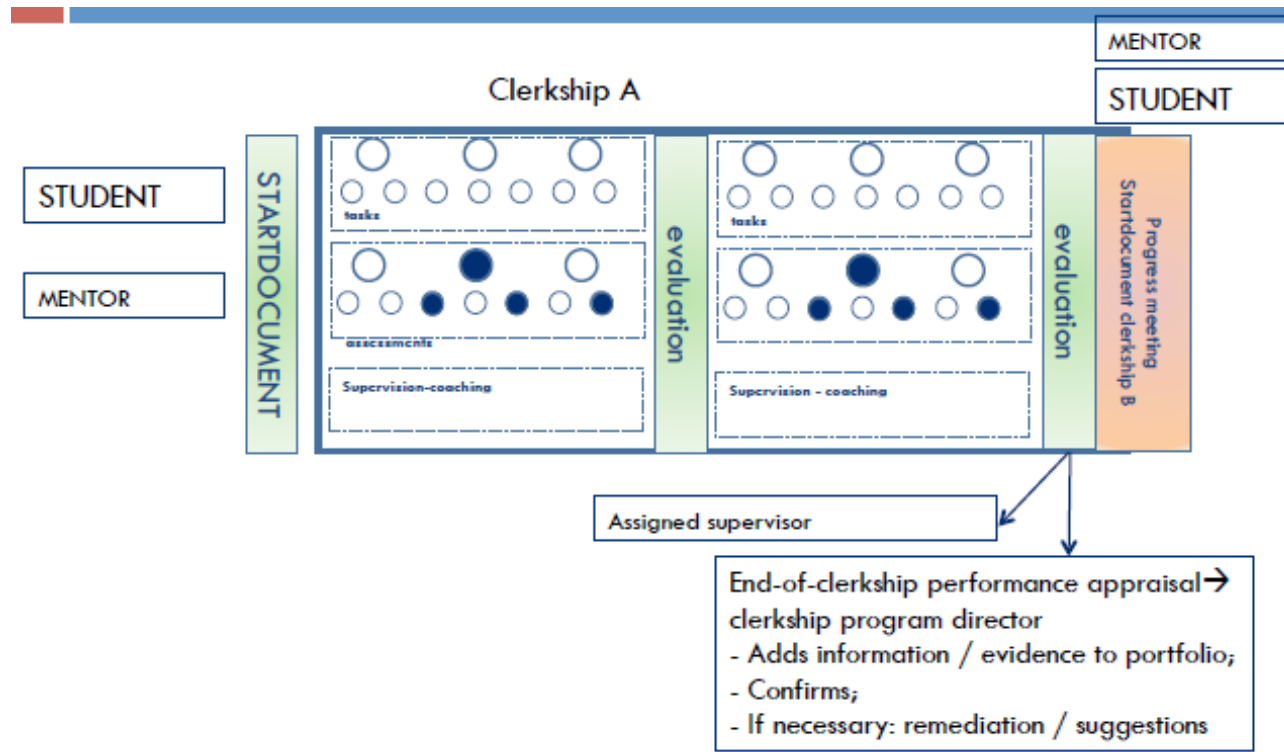
	EPA1	EPA2	EPA3	EPA4	EPA5
Medical expert		++	+		++
Collaborator	+		*	++	
Communicator	+	++			+
Manager		+	++	++	
Health advocate	+		++	+	
Scholar	+				++
Professional	+	+	*		

Assessment

Source: Ten Cate



CBME, EPA's & Programmatic assessment



Neue Rolle der Lehrenden & Lernenden



- Lernende übernehmen Verantwortung für die eigenen Lernprozesse und Lernergebnisse
- Lernende und Lehrende/Coaches verstehen sich als Team
- Coaches stimulieren Reflexion durch Feedback, unterstützen bei der Zielsetzung und Zielerreichung

Entscheidungsfindung im Kollegium

- Kreis der Mentoren und „externe Prüfer“
- Entscheidungen basieren auf Portfolio-Informationen, Mentoren-Beurteilungen und Leistungsstandards
- Es findet ein kritischer Dialog statt (hängt von Klarheit der Informationen ab!)
- Entscheidungen werden begründet, dokumentiert und Empfehlungen werden ausgesprochen



Umgang mit der Subjektivität

- Stichprobengröße („multiple data points“) reduziert zufällige Fehler

- Prozedurale Strategie reduzieren systemische Fehler durch:
 1. Reichhaltige Dokumentation
 2. Involvierung von Studierenden, Patienten, Therapeuten, Experten
 3. Schulung der MentorInnen und Coaches
 4. Kritische Diskussion im Komitee bis zur Erzielung einer gemeinsam getragenen Entscheidung

Umgang mit der „Subjektivität“

Rater-Based Assessments as Social Judgments: Rethinking the Etiology of Rater Errors

Andrea Gingerich, Glenn Regehr, and Kevin W. Eva

Abstract

Background

Measurement errors are a limitation of using rater-based assessments that are commonly attributed to rater errors. Solutions targeting rater subjectivity have been largely unsuccessful.

Method

This critical review examines investigations of rater idiosyncrasy from impression formation literatures to ask new questions for the parallel problem in rater-based assessments.

Results

Raters may form categorical judgments about ratees as part of impression formation. Although categorization can be idiosyncratic, raters tend to consistently construct one of a few possible interpretations of each ratee. If raters naturally form categorical judgments, an assessment system requiring ordinal or interval ratings may inadvertently introduce conversion errors

due to translation techniques unique to each rater.

Conclusions

Potential implications of raters forming differing categorizations of ratees combined with the use of rating scales to collect categorical judgments on measurement outcomes in rater-based assessments are explored.

Seeing the ‘black box’ differently: assessor cognition from three research perspectives

Andrea Gingerich,¹ Jennifer Kogan,² Peter Yeates,³ Marjan Govaerts⁴ & Eric Holmboe⁵

Vielen Dank für Ihre Aufmerksamkeit!

