ANNO DOMINI 1900

EDUCATION

3 Purposes to Sort Out

1. The evaluation of teaching
2. The improvement of teaching
3. Basic research on teaching

How is it best to measure teaching quality?
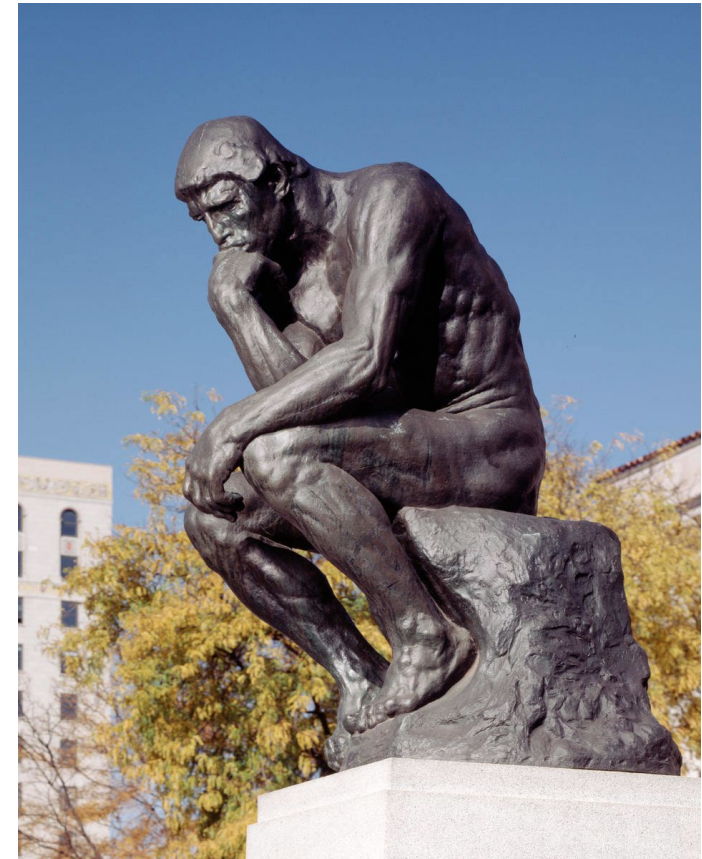
How is it best to measure teaching quality?
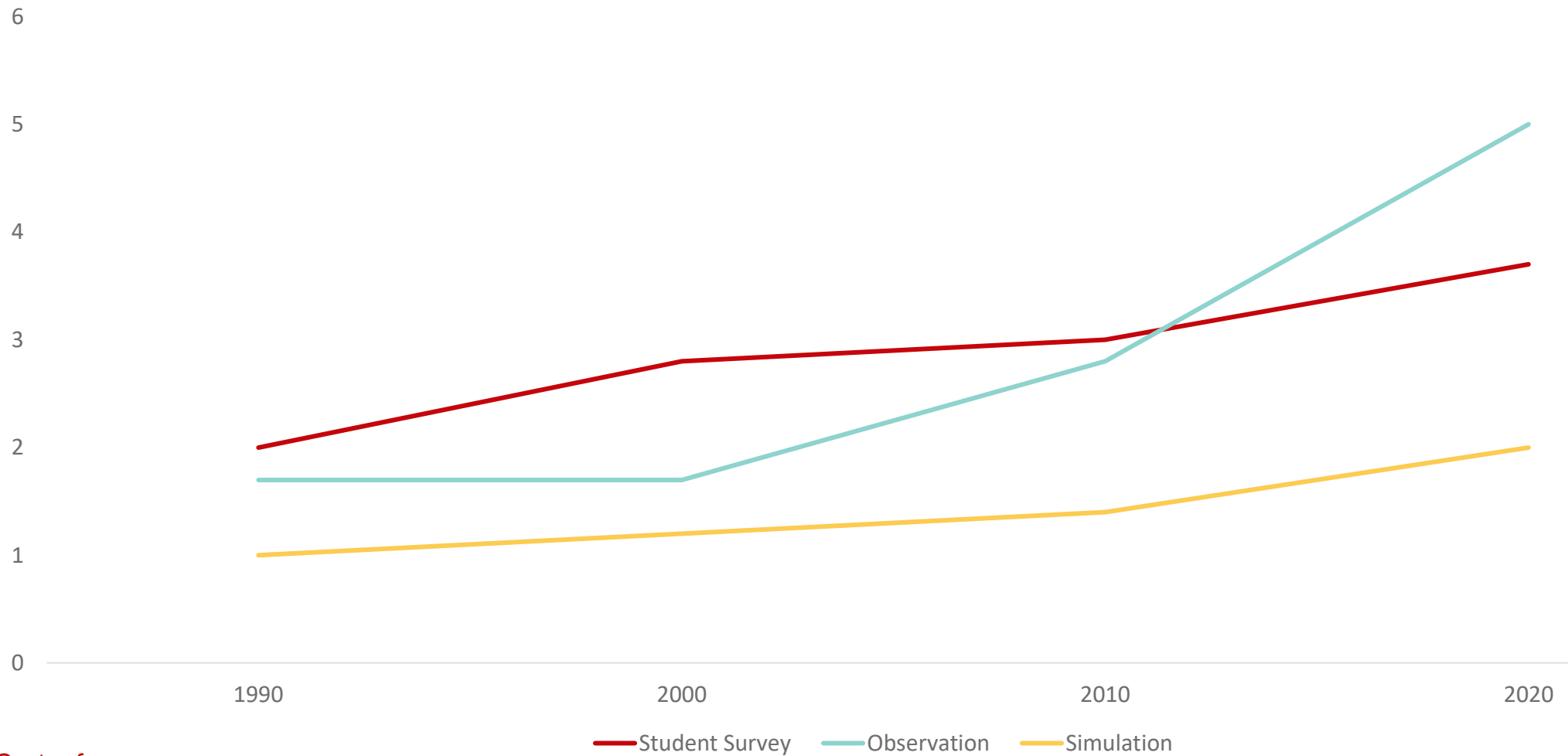
Purpose

Context

Schools

Burden

Definition

Cost

Validity

Community's definition of teaching quality

Reliability

Students

Wisconsin Center for Education Research
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN–MADISON

Increasing research on teaching quality

Talk overview

- Define teaching quality in socio-historical perspective

- Current knowledge about measurement modes

- New measures of teaching quality

- Systematic review of measures of teaching quality

- Enduring tensions of measuring teaching quality

Defining Teaching Quality

Teacher v. Teaching Quality

Imagine two surgeons.

Teacher v. Teaching Quality

Medical care?

Imagine two surgeons. We would like to evaluate them on the quality of their surgical skills using multiple measures. We will use the size of the scar, the rate of infection, the quality of pain management, and patient satisfaction as our measures of the quality of their surgical skills. One is in Miami, Florida, the other in Moshe, Tanzania. Both must remove a benign tumor from a 53 year old man's abdomen. The surgeon in Miami has a #10 blade steel scalpel that is designed for cutting muscle and skin. The surgeon in Moshe has a well sharpened utility knife that is used for a range of surgical purposes. The excision in Miami will occur in a sterile operating room with no external windows, fans and filters to circulate and clean the air, an anesthesiologist, and a surgical nurse. The excision in Moshe will occur in a clean operating room washed with well water and bleach, windows opened a crack to allow the breeze to circulate the stiflingly hot air, no fans or filters, and a nurse borrowed from the pediatrics unit because she was the only available help.

Gitomer & Bell, 2013

Attribution

The purpose of psychology

To predict, given the stimulus, what reaction will take place; or, given the reaction, state what the situation or stimulus is that has caused the reaction

Watson, 1930, p. 11

The shift in the social sciences

Behaviorism

This view of teaching relates to transmission: it is the role of the teacher to transmit knowledge to students…It is basically a transfer of knowledge from the teacher to the student and the purpose is to "get" the student to do something; either a set of skills, or being able to answer a knowledge test with some accuracy. Within the classroom, the teacher, text-book, and or other media are the experts. The content to be mastered by the students resides in the teacher's talk, the written word, or a media presentation.

Richardson (1996)

Situated and sociocultural

The teacher's role is to monitor students' …interactions … attend to the conceptual ideas with which students are working and the ways in which their thinking is similar and different from that of their classmates."

National Research Council (2005)

"…teachers' instructional practices are not merely influenced by but are partially constituted by the materials and resources that the teachers use in their classroom practice, the institutional constraints that they attempt to satisfy, and the formal and informal sources of assistance on which they draw.

Cobb, et al. (2009)

The shift in the social sciences

Behaviorism

This view of teaching relates to transmission: it is the role of the teacher to transmit knowledge to students…It is basically a transfer of knowledge from the teacher to the student and the purpose is to "get" the student to do something; either a set of skills, or being able to answer a knowledge test with some accuracy. Within the classroom, the teacher, text-book, and or other media are the experts. The content to be mastered by the students resides in the teacher's talk, the written word, or a media presentation.

Richardson (1996)

Situated and sociocultural

The teacher's role is to monitor students' …interactions … attend to the conceptual ideas with which students are working and the ways in which their thinking is similar and different from that of their classmates."

National Research Council (2005)

"…teachers' instructional practices are not merely influenced by but are partially constituted by the materials and resources that the teachers use in their classroom practice, the institutional constraints that they attempt to satisfy, and the formal and informal sources of assistance on which they draw.

Cobb, et al. (2009)

The shift in the social sciences

Behaviorism

This view of teaching relates to transmission: it is the role of the teacher to transmit knowledge to students…It is basically a transfer of knowledge from the teacher to the student and the purpose is to "get" the student to do something; either a set of skills, or being able to answer a knowledge test with some accuracy. Within the classroom, the teacher, text-book, and or other media are the experts. The content to be mastered by the students resides in the teacher's talk, the written word, or a media presentation.

Richardson (1996)

Situated and sociocultural

The teacher's role is to monitor students' …interactions … attend to the conceptual ideas with which students are working and the ways in which their thinking is similar and different from that of their classmates."

National Research Council (2005)

"…teachers' instructional practices are not merely influenced by but are partially constituted by the materials and resources that the teachers use in their classroom practice, the institutional constraints that they attempt to satisfy, and the formal and informal sources of assistance on which they draw.

Cobb, et al. (2009)

Definitions of teaching quality should attend to....

- The goals of schooling
- The social context – in and out of school
- The teacher's role
- The student's role
- Resources
- Individuals and their specific selves

Wisconsin Center for
Education Research
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN–MADISON

Attribution

Attribution → Theory

Attribution → Theory → Specific Terms

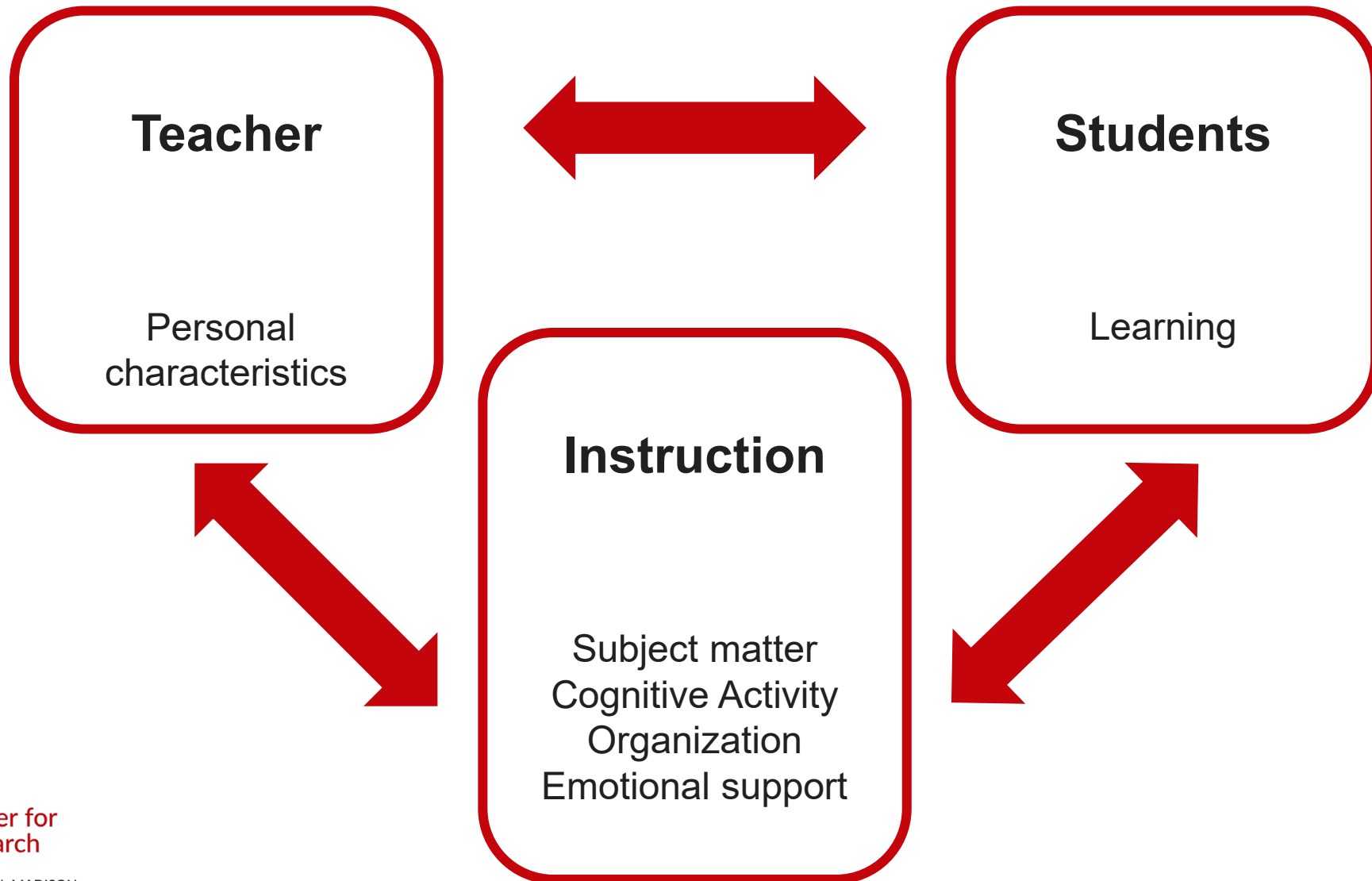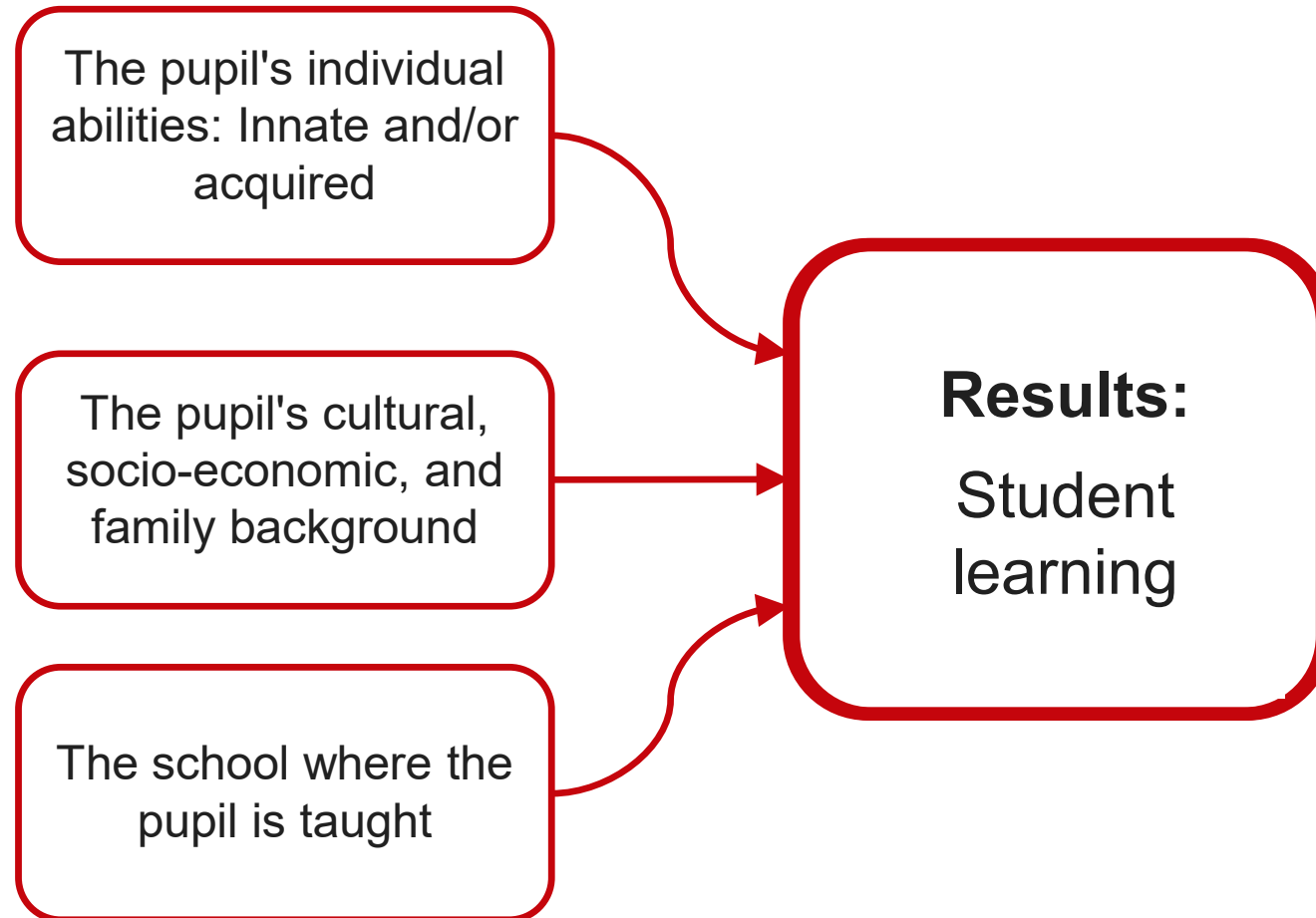|	Teacher quality	Teaching quality	School effectiveness
ERIC Definition	*No teacher quality* Teacher effectiveness: Degree to which teachers are successful in satisfying their objectives, obligations, or functions	*Teaching quality: Only used from 1966 to 1980 – Use teacher effectiveness* Instructional effectiveness: Degree to which instructional materials or programs are successful in accomplishing their objectives	School effectiveness: Degree to which schools are successful in accomplishing their educational objectives or fulfilling their administrative, instructional, or service functions
Literature Definition	The personal characteristics (knowledge, beliefs, capabilities) a teacher uses to engage students in learning. (c.f., Fauth et al., 2019)	Instructional quality: "Observable characteristics of classroom instruction that are orchestrated <u>by</u> teachers and <u>goes along with</u> desirable development of students' learning outcomes in a theoretically plausible way, supported by empirical evidence." (Mu, et al., 2022)	"…school effectiveness, defined as the causal effect of attending a particular school or set of similar schools (like charter schools) on student outcomes." (Angrist, et al., 2022)
Model	1	2	3
Attribution	Courtney has X level of Y skill. Mariana caused the students to learn Z.	Courtney enacts X level of Y skill in W context, with Q students. Classroom interactions caused the students to learn Z.	Rock Park Primary School supports student learning of Z through Courtney and Mariana's Y skills.

Teacher quality – Model 1

Teacher

Personal characteristics

Instruction

Subject matter
Cognitive Activity
Organization
Emotional support

Student

Learning
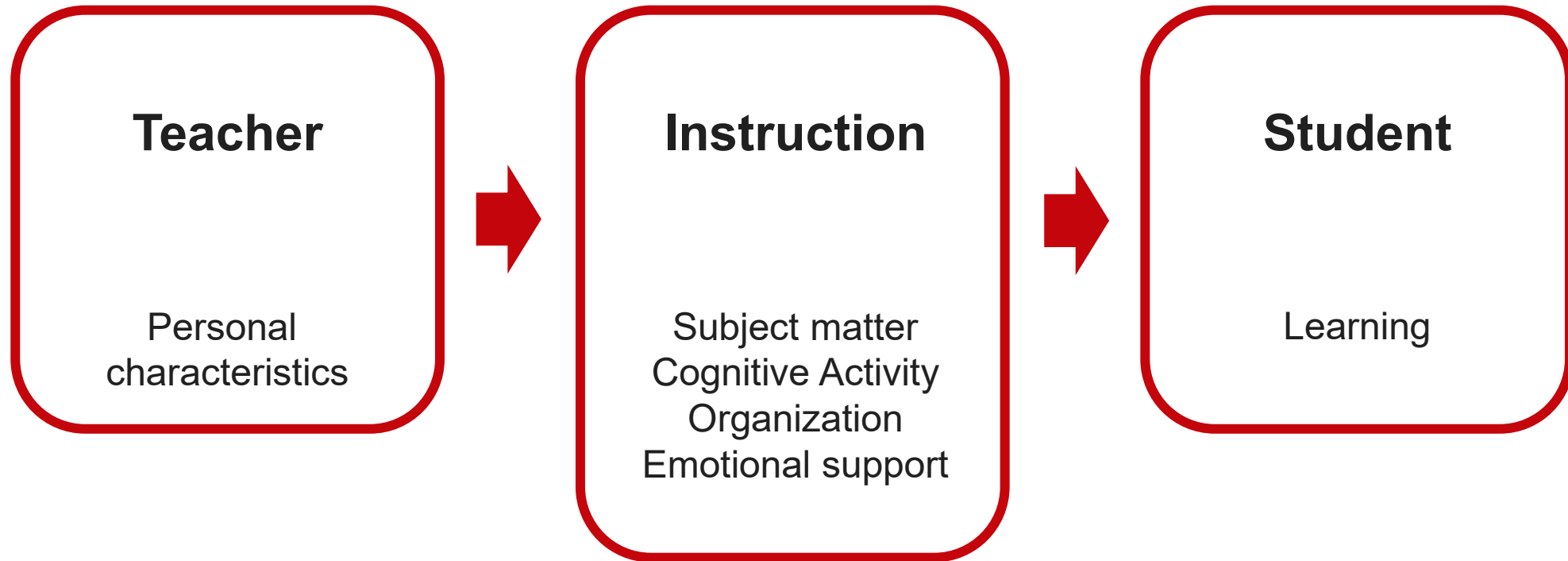
Teacher quality – Model 1

Teaching quality – Model 2

Teacher

Personal characteristics

Students

Learning

Instruction

Subject matter
Cognitive Activity
Organization
Emotional support
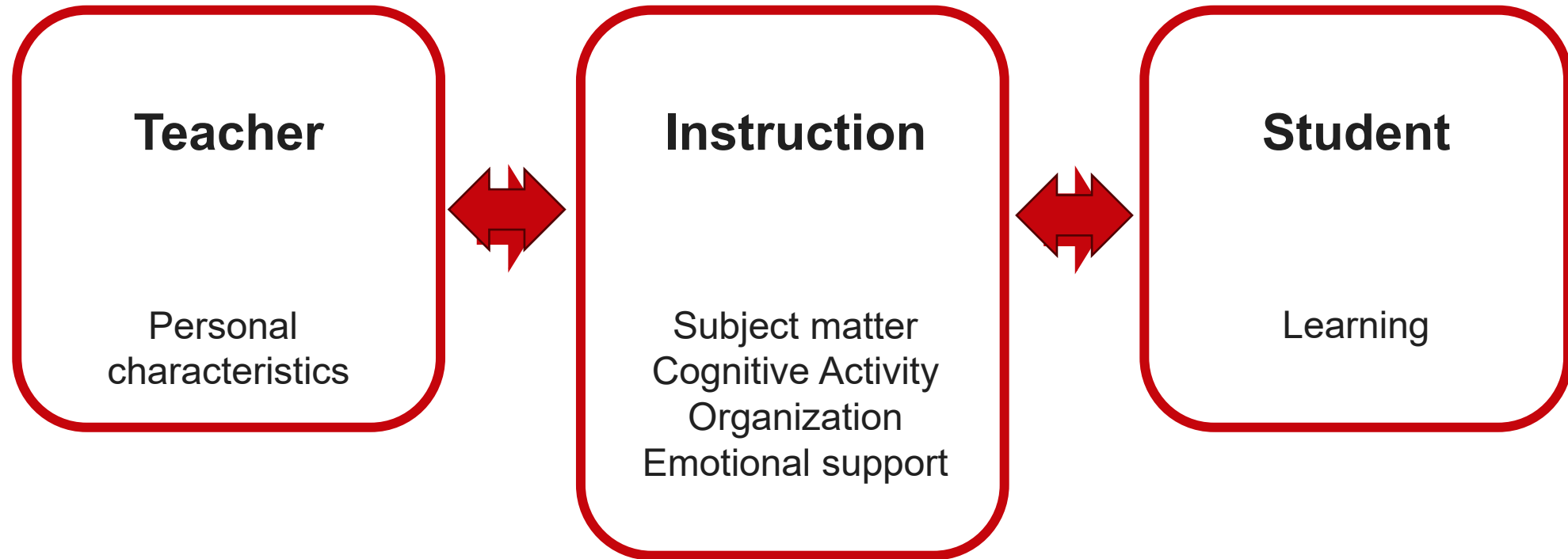
School effectiveness – Model 3

The pupil's individual abilities: Innate and/or acquired

The pupil's cultural, socio-economic, and family background

The school where the pupil is taught

Results:
Student learning

Nordenbo et al. (2009)

	Teacher quality	Teaching quality	School effectiveness
ERIC Definition	*No teacher quality* Teacher effectiveness: Degree to which teachers are successful in satisfying their objectives, obligations, or functions	*Teaching quality: Only used from 1966 to 1980 – Use teacher effectiveness* Instructional effectiveness: Degree to which instructional materials or programs are successful in accomplishing their objectives	School effectiveness: Degree to which schools are successful in accomplishing their educational objectives or fulfilling their administrative, instructional, or service functions
Literature Definition	The personal characteristics (knowledge, beliefs, capabilities) a teacher uses to engage students in learning. (c.f., Fauth et al., 2019)	Instructional quality: "Observable characteristics of classroom instruction that are orchestrated <u>by</u> teachers and <u>goes along with</u> desirable development of students' learning outcomes in a theoretically plausible way, supported by empirical evidence." (Mu, et al., 2022)	"…school effectiveness, defined as the causal effect of attending a particular school or set of similar schools (like charter schools) on student outcomes." (Angrist, et al., 2022)
Model	1	2	3
Attribution	Courtney has X level of Y skill. Mariana caused the students to learn Z.	Courtney enacts X level of Y skill in W context, with Q students. Classroom interactions caused the students to learn Z.	Rock Park Primary School supports student learning of Z through Courtney and Mariana's Y skills.

# How is it best to measure teaching quality?
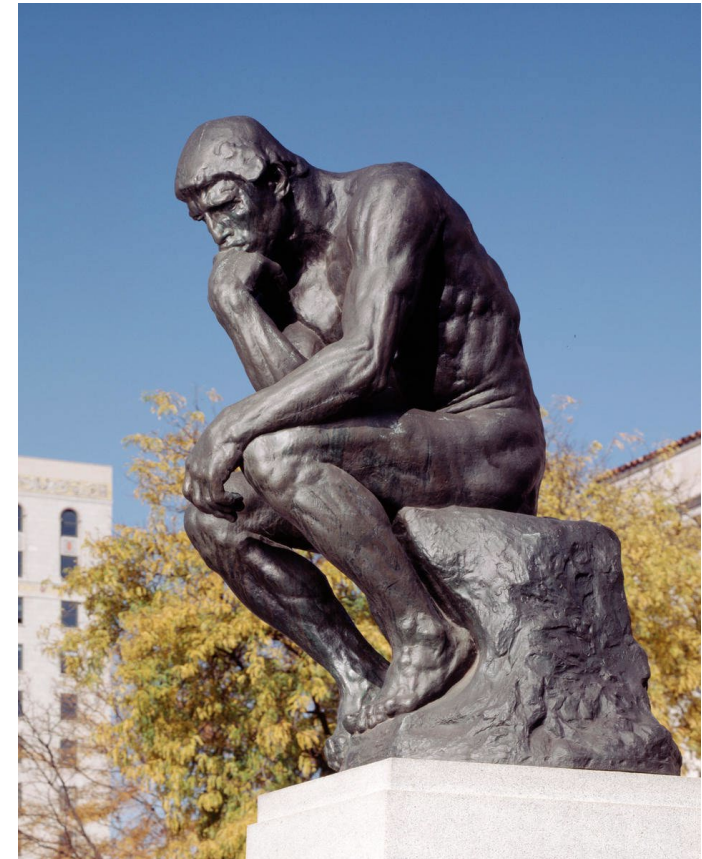
Constructs that can be measured

- Teacher characteristics (knowledge, beliefs, etc.)
- Student characteristics (knowledge, beliefs, capabilities, etc.)
- Discourse (teacher, student, or classroom)
- Embodied knowing and doing (movement, gestures, eye gaze, tone of voice)
- Tools (representations, artefacts, texts, lesson plans)

Interactions

The contextual factors, constructs, and measures associated with teaching quality
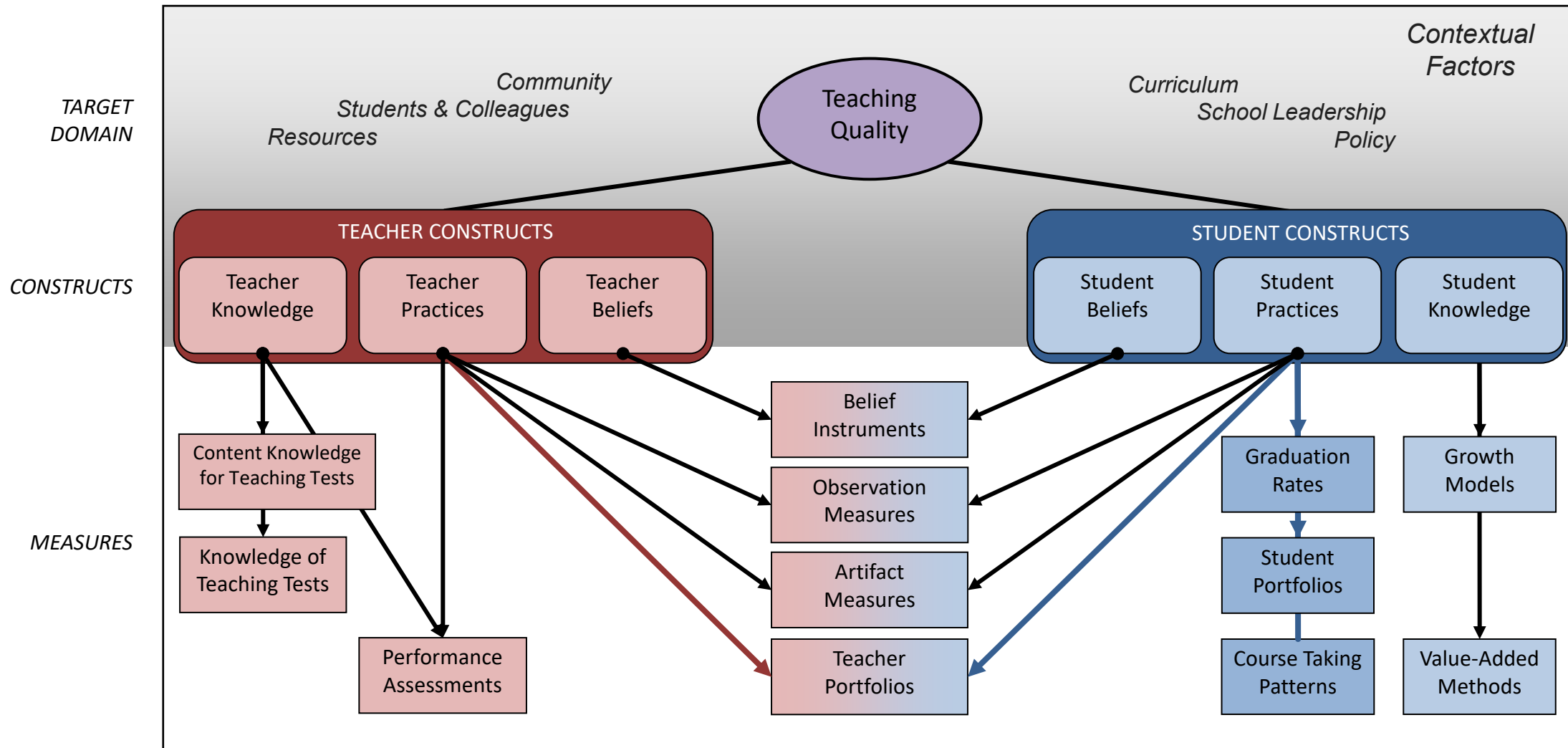
Bell, Gitomer, McCaffrey, Hamre, Pianta & Qi (2012)

Literature review findings

- Definitions of teaching quality are not clearly specified

- Multiple terms are used for the same (and different) things

- Measurement modes are differentially suited to measure different aspects of teaching quality

Current Knowledge of Teaching Quality Measures

Measurement modes in the review

Student survey

Teacher survey

Observation

Artefacts

Simulation

Automated scoring
of video/audio

Biological/physiological
measures

Wisconsin Center for
Education Research
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN–MADISON

What is in a "mode"

- Perspective on classroom interactions
  - Teacher participant
  - Student participant
  - 3rd party (rater)

- Medium – type of evidence
  - Pencil and paper questionnaire – perceptions
  - Artefacts from the lesson – criterion-based judgements
  - Recording of the lesson or simulation – criterion-based judgements

- Aspect of classroom interactions
  - Feedback
  - Disruptions
  - Frequency of topic covered, etc.

- Relationship to the measurement purpose
  - Social relationships (student questionnaire → improve teaching)
  - Societal values and beliefs (trust, objectivity, etc.)

Each mode has affordances and constraints

Affordances of Familiar and New(ish) Measures of Teaching Quality

Mode	Affordance
Student survey	• Important stakeholder in a classroom • Can better interpret how students interpret teachers' actions
Teacher survey	• Important stakeholder in a classroom • Can infer teachers' internal state • Allows for data collection that is challenging to get through other modes
Observation	• High face validity • Can rewatch and pause, if done by video • Can have multiple quality frameworks applied
Artefacts	• Supports inferences about each student (not the whole class) • Can infer what students are cognitively engaged in, different from videos and questionnaires • Can be tightly tied to the subject matter
Simulations	• Can be carried out lower and higher cost • Does not do harm to students which novices are learning • Can be engineered to study the development of competencies • Evidence they are impactful on professional learning (Chernikova, 2020)
Automated scoring	• Allows observation and real-time data be "objectively" coded • Allows data to be efficiently coded, speeding up and decreasing the costs of research
Bio/physiological	• Allows for understanding the embodied nature of teaching and learning • Allows us to study new constructs

The Review Team

Systematic Review

Review research questions

- How reliable are the respective measurement approaches?
- What is the strength of the validity evidence linking teaching quality and student outcomes in each mode?
- What are the practical considerations and costs associated with each measurement mode?

Definitions shaping the review

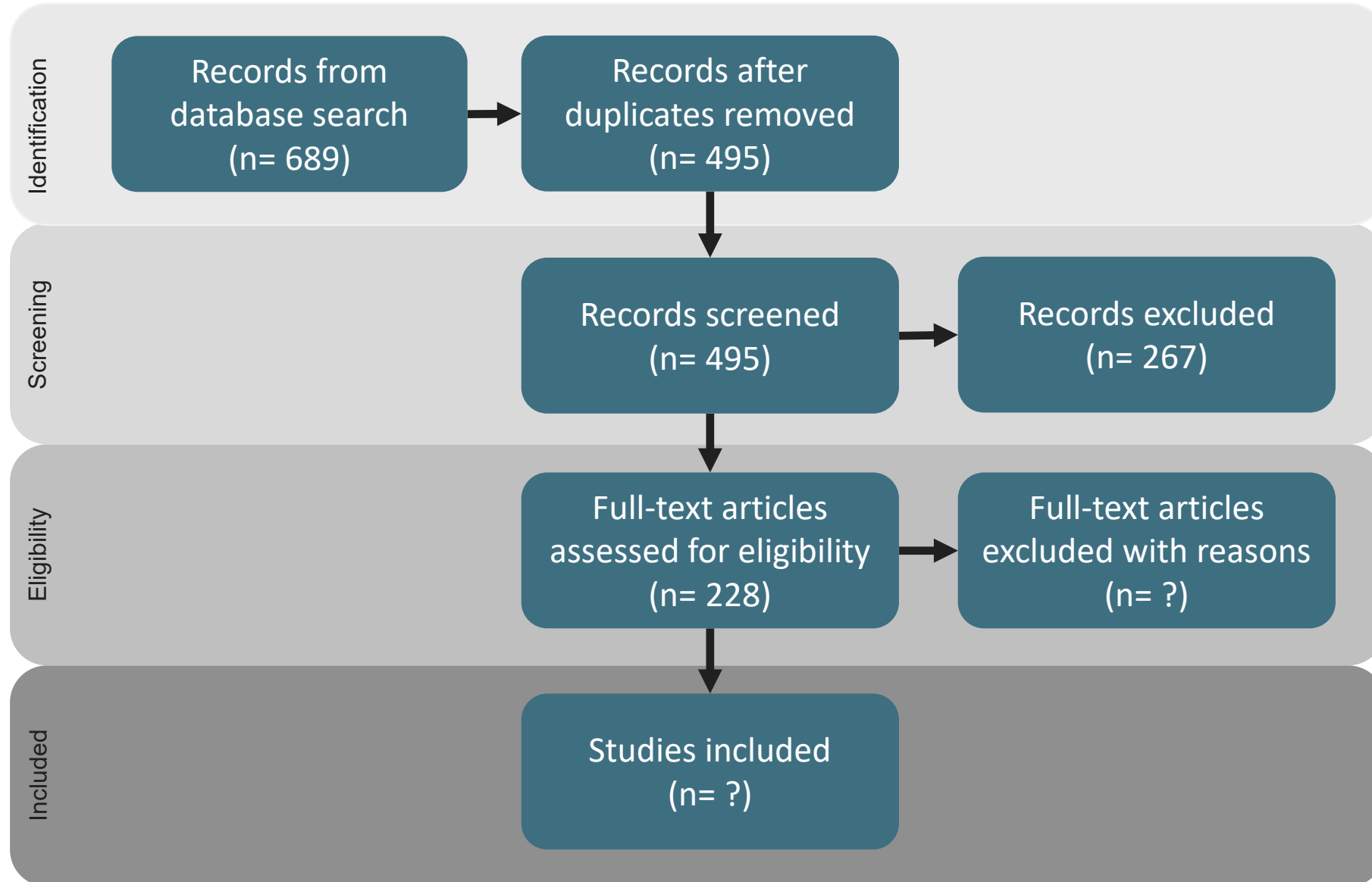| Reliability	For scales and/or instruments: Internal consistency between items designed to measure the same underlying construct For human rated measures: Agreement between raters
Validity	"…*validity*… extent to which theory and evidence supports or refutes proposed and actual or enacted test score interpretations, uses, and consequences, and *validation* as the process of gathering, evaluating, and synthesizing such evidence. " (Marion & Shepard, *forthcoming*)
Usability and Cost	The degree to which the mode is able to be used with a moderate level of effort affordably by researchers to study teaching quality at scale

Search criteria

- Must be around teaching quality (papers should not include ONLY psychological constructs such as efficacy, motivation, etc.)

- Articles must include at least one measure of teaching quality for at least one aspect of teaching quality

- English language

- Grade levels: K-12, no preK

- Years 2008-2022

- Peer-reviewed journals

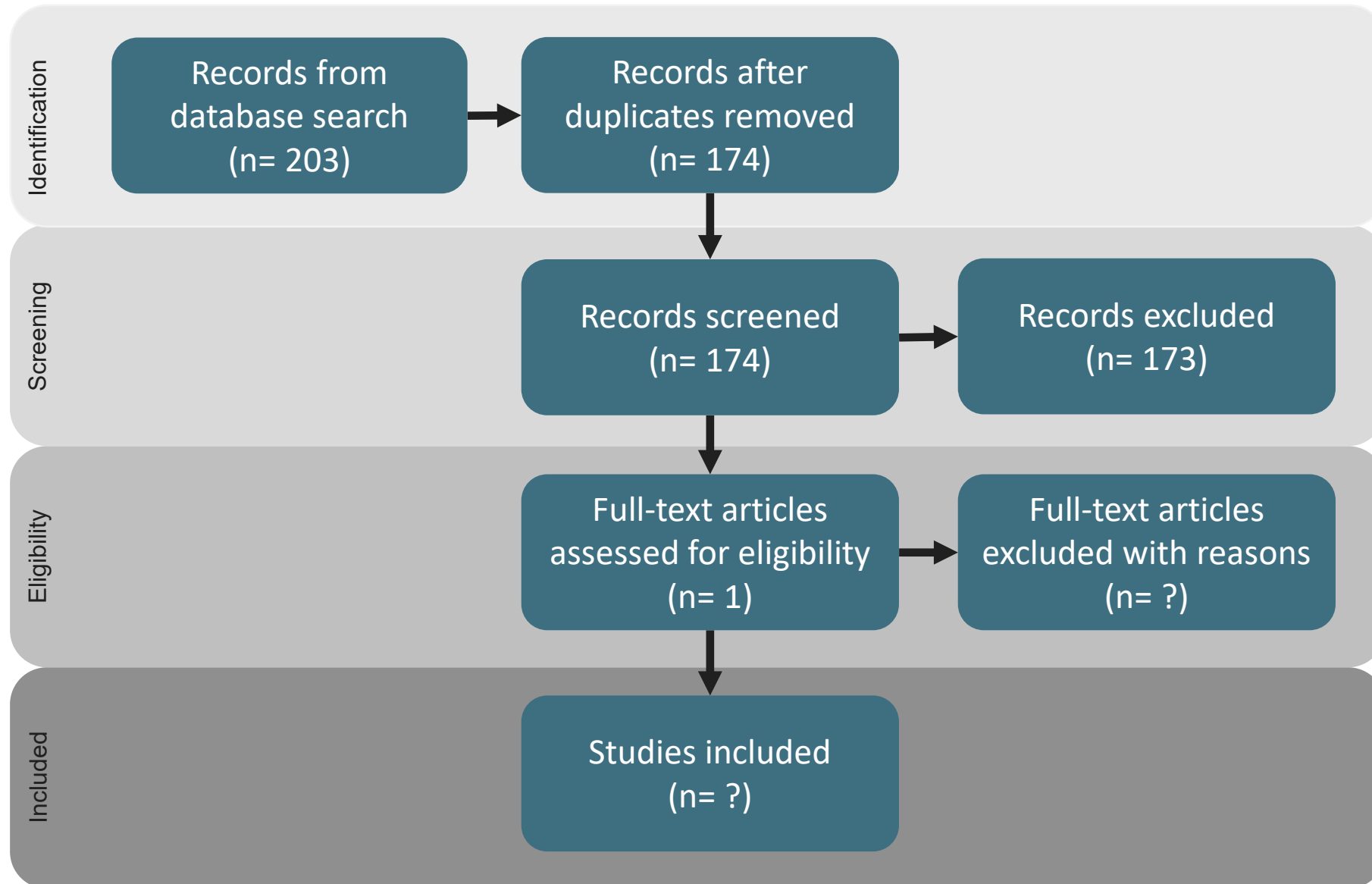Web of Science, Psych Info, and Education Research Complete & ERIC (via Ebsco)

Overall

Identification

Records from database search (n= 2590) → Records after duplicates removed (n= 1985)

Screening

Records screened (n= 1985) → Records excluded (n= 1287)

Eligibility

Full-text articles assessed for eligibility (n= 698) → Full-text articles excluded with reasons (n= ?)

Included

Studies included (n= ?)

Observations

Identification

Records from database search (n= 689) → Records after duplicates removed (n= 495)

Screening

Records screened (n= 495) → Records excluded (n= 267)

Eligibility

Full-text articles assessed for eligibility (n= 228) → Full-text articles excluded with reasons (n= ?)

Included

Studies included (n= ?)

Simulation

Why were some articles not summarized?

a. It was a review article
b. It was not a measure of teaching quality
c. No instrument or tool was described and/or used
d. There were not enough items (fewer than 3 items)
e. The study was qualitative and could not be replicated.

Literature review findings

- Definitions of teaching quality are not clearly specified

- Multiple terms are used for the same (and different) things

- Measurement modes are differentially suited to measure different aspects of teaching quality

- We know much less about measuring teaching quality with the newer modes
  - Observations, teacher surveys and student surveys searches pulled up many more relevant articles than the other modes
  - Newer modes require more nomination and there is less overlap in reference lists than more established modes

- Newer modes are measuring narrower aspects of teaching quality

Provisional findings for selected modes

Simulation

Reliability	• Rubric-based; rated by humans • Have few scales/items to combine • Challenging to distribute scores for variation in scores
Validity	• Focused in the primary grades and classroom management • Used formatively; few replicated findings or study designs • Have only produced a small numbers of instruments that can be used by others • Teachers enjoy them and find them authentic • Trade off: capturing construct of teaching quality and reliability
Usability and Cost	• Expensive to use virtual and mixed reality (VR and MR) • Trained interactors may be prohibitively expensive for some types of tasks and simulations (e.g., MR). • Expensive to develop scenarios and train interactors (MR) • Likely to be best used in University settings

Artefacts

Reliability	• With standardized curricula little variation in scores undermines reliability • Rubric based; rated by humans • Challenge to reliably rate varied data (or a lot of data)
Validity	• If artefacts vary across contexts, inferences may have weak validity evidence • Attribution can be problematic, depending on artefacts collected and attributions being made
Usability and Cost	• Most research is in the US • Increasingly easy to scan on site with phone-based scanning technology and electronic assignments. • Burdensome to teacher and researcher – many data records to link and track • Researchers use deep knowledge of local classroom and curricular contexts

Four Enduring Tensions and Opportunities for Lines of Research

Potential research trajectories

- Biological/physiological, automated scoring, simulation
  - Develop collaborations between teaching scholars and others
  - Develop clear use cases
  - Match mode with aspect of teaching quality

- Across mode
  - Developing knowledge about how to create rubrics/scales that reflect (meaningful) variation in scores
    - In simulation – will need to occur when scenarios are developed
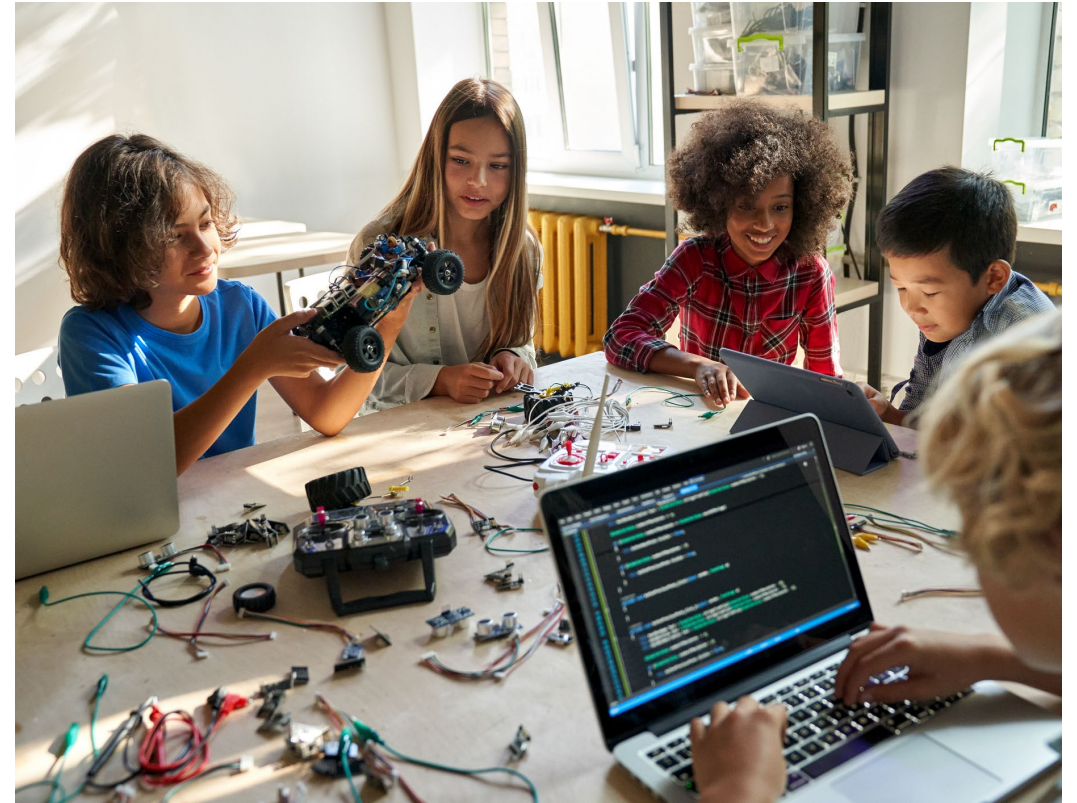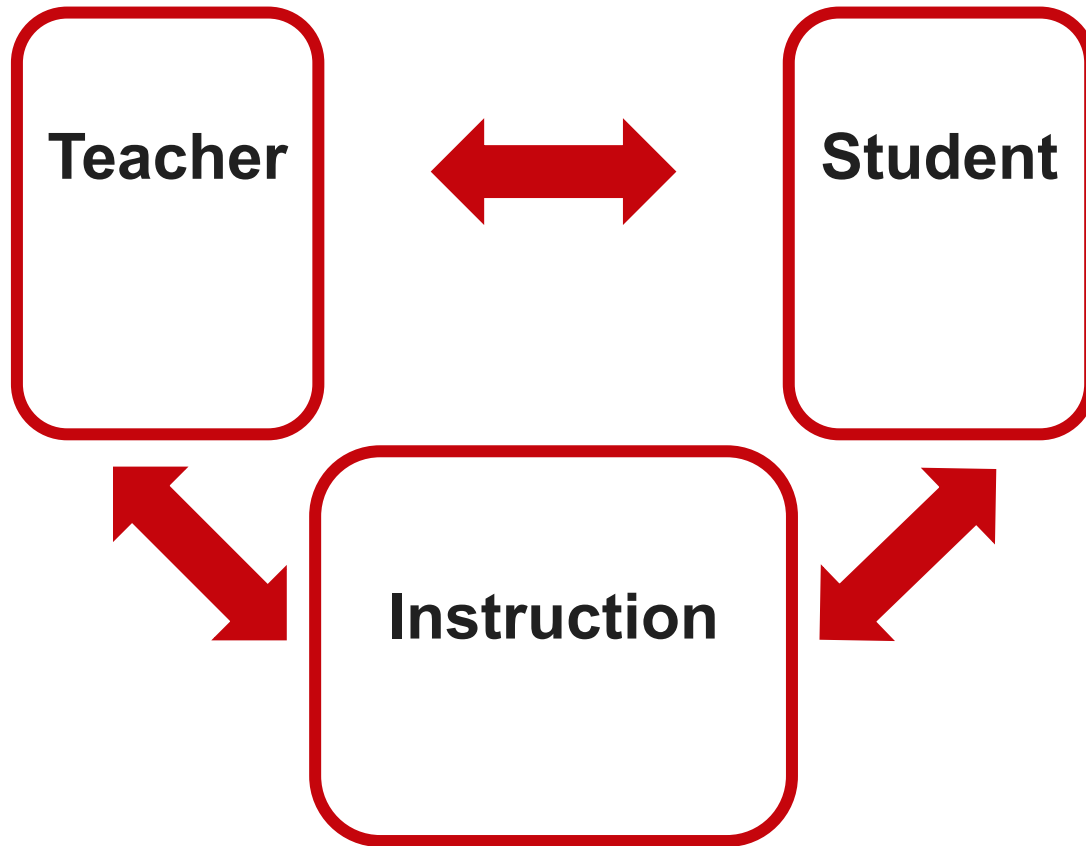   - In automated scoring – meaningful metrics for parsing discourse

How best to account for purpose

Facing constraints to acknowledge "It is good enough for what it's for"

How to account empirically for our theories of teaching and learning

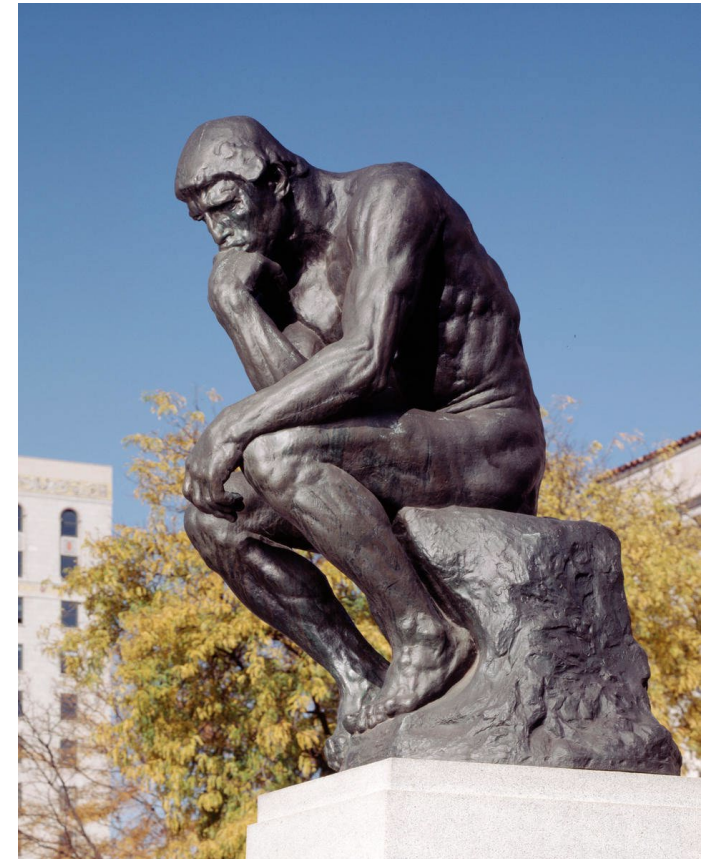**Teacher** ⟷ **Student**

Instruction

Keeping human judgement and wisdom in our assessment systems

How is it best to measure teaching quality?

- Purpose

- Teaching quality definition

- Attribution via models

- Mode affordances and constraints

- Technical quality

- Constraints of social and research contexts

- Human judgement in complex social systems

References

Angrist, J., Hull, P., & Walters, C. R. (2022). Methods for measuring school effectiveness.

Bell, C. A., Gitomer, D. H., McCaffrey, D., Hamre, B., Pianta, R., Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2–3), 62-87. https://doi.org/10.1080/10627197.2012.715014

Cobb, P., Zhao, Q., & Dean, C. (2009). Conducting design experiments to support teachers' learning: A reflection from the field. *The Journal of the Learning Sciences*, *18*(2), 165-199.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, *90*(4), 499-541.

Fauth, B., Decristan, J., Decker, A. T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and teacher education*, *86*, 102882.

Mu, J., Bayrak, A., & Ufer, S. (2022, October). Conceptualizing and measuring instructional quality in mathematics education: A systematic literature review. In *Frontiers in Education* (Vol. 7, p. 994739). Frontiers.

Bransford, J. D., & Donovan, M. S. (Eds.). (2005). *How students learn: History, mathematics, and science in the classroom*. National Academies Press.

Nordenbo, S. E., Holm, A., Elstad, E., Scheerens, J., Larsen, M. S., Uljens, M., ... & Hauge, T. E. (2005). Input, Process, and Learning in primary and lower secondary. *Educational Research Journal*, *17*(3), 203-217.

Richardson, V. (1996). From Behaviorism To Constructivism In Teacher Education1. *Teacher Education and Special Education*, *19*(3), 263-271.

Watson, J. B. (1930). *Behaviorism* (Rev. ed.). W W Norton & Co.