

Learn and let learn: Why there is no shortcut to a valid measurement of learning

Jens F Beckmann
Durham University
UK j.beckmann@durham.ac.uk

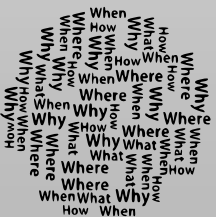
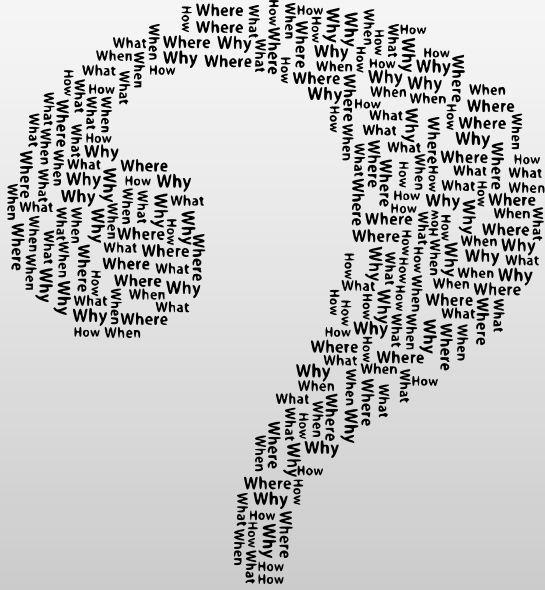
ZIB Academy 2026
Learning & Assessment



19 – 23 July 2027
issid2027.org

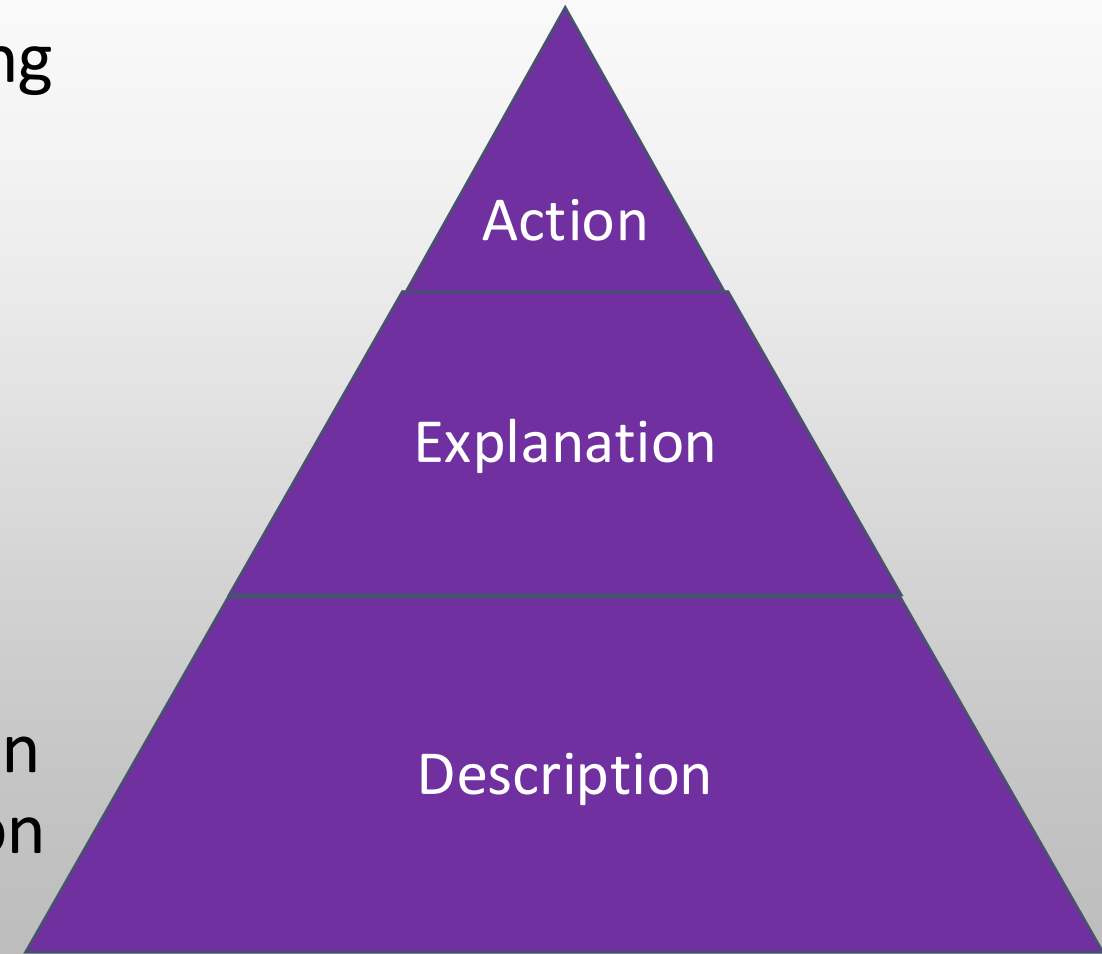
Learning & Assessment

Why do you care?



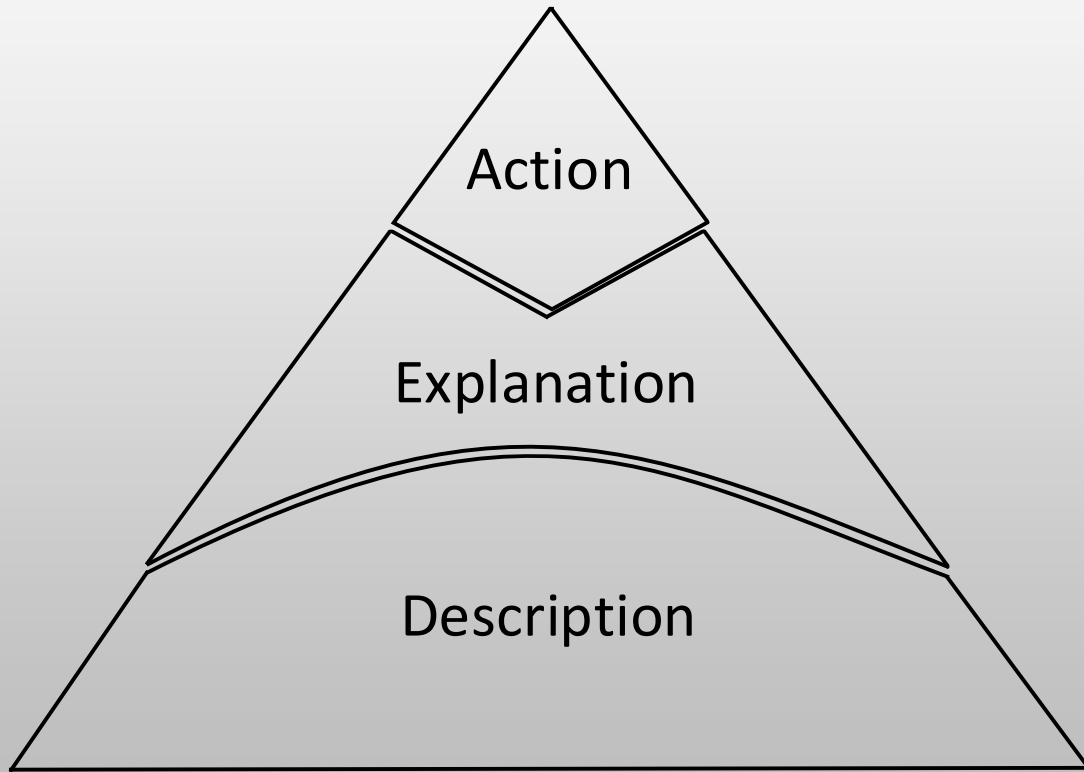
Why do we care about learning?

- ... to optimise / improve / facilitate ... learning
- **Keep in mind:**
 - **Interventions** (i.e., actions) should be **based on an understanding** of the conditions and processes **causal** to learning
 - an **understanding** of the conditions and processes involved in learning **requires** an adequate **description** of the phenomenon we call learning.



The hierarchy of research objectives

Hierarchy of Research Objectives



Meaningful **prescription of actions** depend on an understanding of the causal mechanisms behind the manifestation of the phenomenon in question

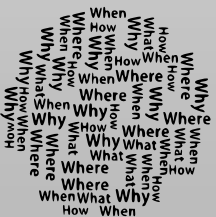
Explanations (i.e., answering the **Why**-questions) is necessary for devising meaningful action

A thorough description of the phenomenon of interest (i.e., answering the **What**-question) is the necessary foundation for all research

Learning & Assessment

Why do you care?

Who are you?



Research

Consideration of imminent applicability?

Quest for fundamental understanding?

	no	yes
yes	 Pure basic research Bohr	 Use-inspired basic research Pasteur
no	 Linnaeus	 Pure applied research Edison

What is Education (in the context of research)?

- It is not an **academic discipline**
 - *It therefore does not lend itself to being a form of a to be adopted “academic identity”*
- It is an **applied field of study**, informed by a multitude of disciplines
 - *It therefore requires a multi-disciplinary perspective*

Psychology: multi-paradigmatic scientific discipline

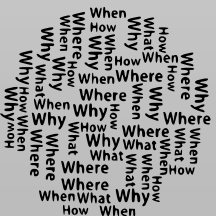
Education: multi-disciplinary applied field of study

Learning & Assessment

Why do you care?

Who are you?

What do you do?



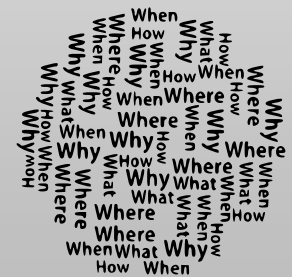
What is learning?

Is it ...

- an **Activity**: “Learning is what people do”
 - an **Outcome**: “Learning is what people have done”
 - a **Process**: “Learning is change”
- ... change in what, exactly?

behaviour

knowledge, skills, ability, attitude, motivation, personality, emotions ...





Learning as a Process

What is a process?

- a transition between (at least) two states

How would we know?

- observation, measurement of change

How do we do that?

- assessment as an attempt to **describe** observations (with a sufficient degree of systematicity)

But what about explanation?



manifest
(observable)

Description



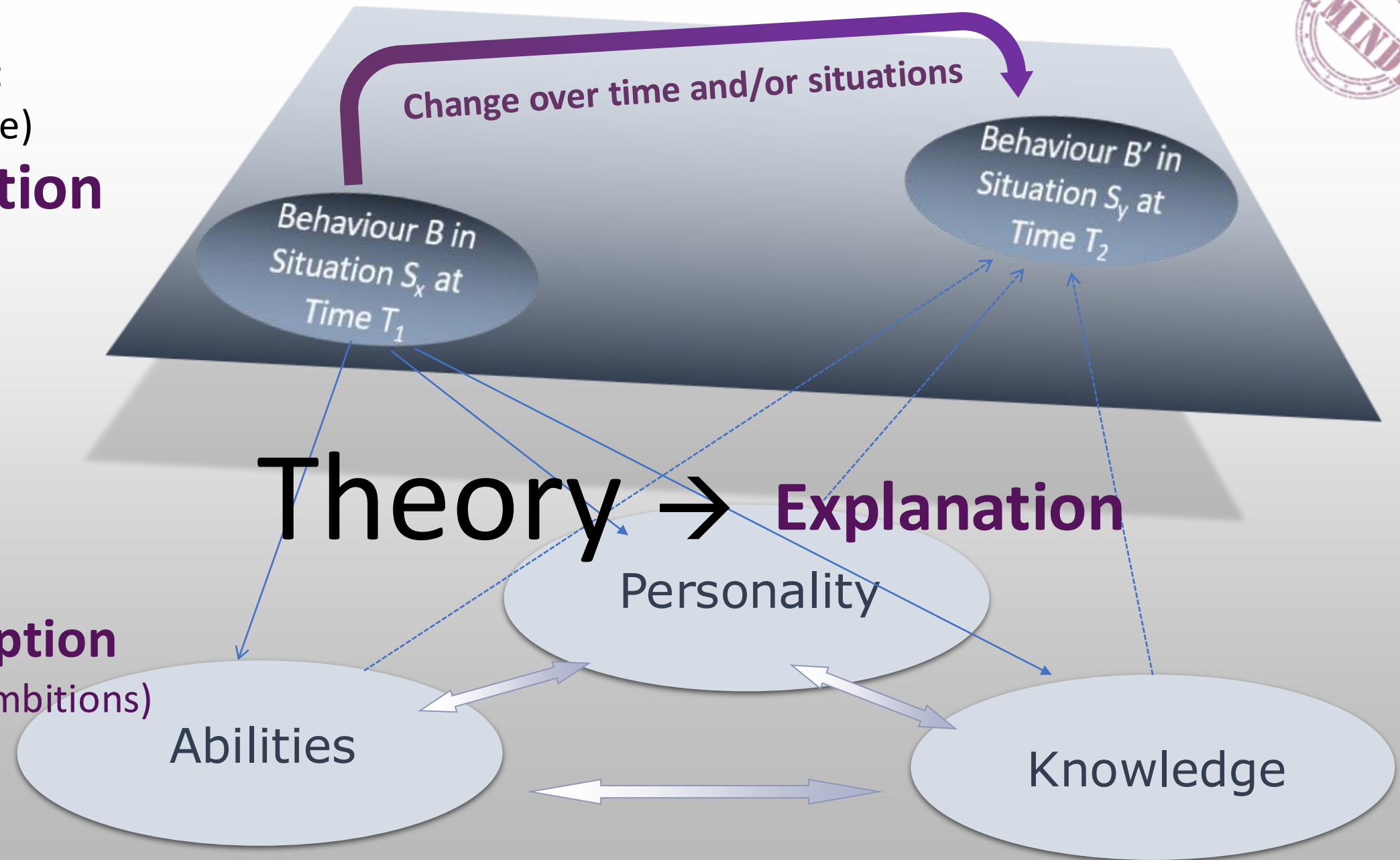
But what does this mean?



manifest
(observable)
Description



(Still) **Description**
(explanatory ambitions)
latent
(constructs)



Understanding Learning ...



... as an abstract concept (construct)

- **Observation:** “change in behaviour over time” ← **description**
- **Interpretation:** by making reference to latent variables or constructs: e.g. “learning in form of an increase in knowledge” ← **explanation**

➤ Link between Observation and Interpretation: **Theories**

- Theories as speculative conceptual explanations, specify expectations (predictions) regarding future behaviour
- **Theories are causal explanations**

➤ **But** not all changes in behaviour are indicative of learning

- e.g. maturation, development, (short-term) changes caused by illness, drug use etc. is not learning

➤ Different perspectives (e.g., **paradigms** in psychology) carry different interpretations for observed changes in thoughts, feelings, attitudes, and behaviour

Theories



- ... are a systematic way of **organising** and **explaining** observations
- ... are a **set of propositions** that *describe* and *aim to explain* phenomena of interest
- ... help us **make sense** of phenomena observed in the (educational) world
- not to be **confused** with speculations, ideas, hunches, a point of view ...

To, C.-Y. (2000) The scientific merit of the social sciences: Implications for research and application. Stoke on Trent: Trentham Books.

What is a (scientific) theory?

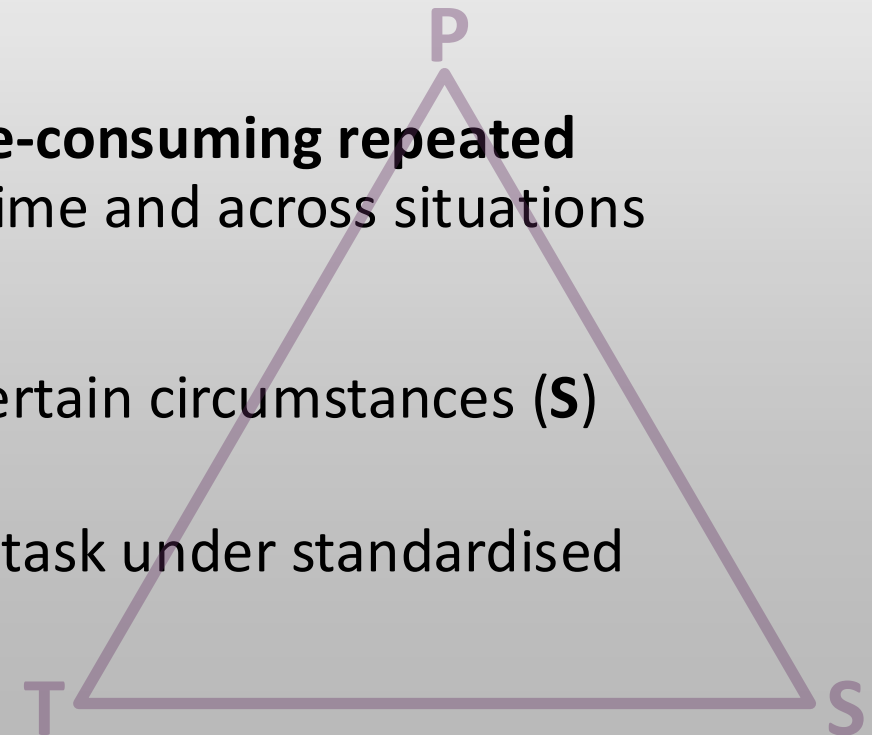
- A **logically integrated set** of propositions about latent (not directly observable) mechanisms, processes, and variables, and their causal relations to each other and to manifest (observable) variables.
- ... provide **explanations** that are formulated at an abstract (and therefore generalisable) level and are **testable** through empirical research.
- help us **make sense** of phenomena observed in the (educational) world.

eg Oberauer & Lewandowski 2019, p. 1597; Van Lange 2013, p. 41



More Questions

- **What is learning?**
 - as concept it is a label (or **descriptor**) for **some form of (change in) behaviour**
- **What is the assessment of learning?**
 - as an activity it is an attempt to shortcut the **time-consuming repeated observation of said behaviour** (efficiency) over time and across situations
- **How can this be done?**
 - We ask a Person (**P**) to do something (**T**) under certain circumstances (**S**)
 - Person-Task-Situation framework
 - such demand constellation (a person performs a task under standardised situations) might be called a (psychometric) test



Where is ?

Person

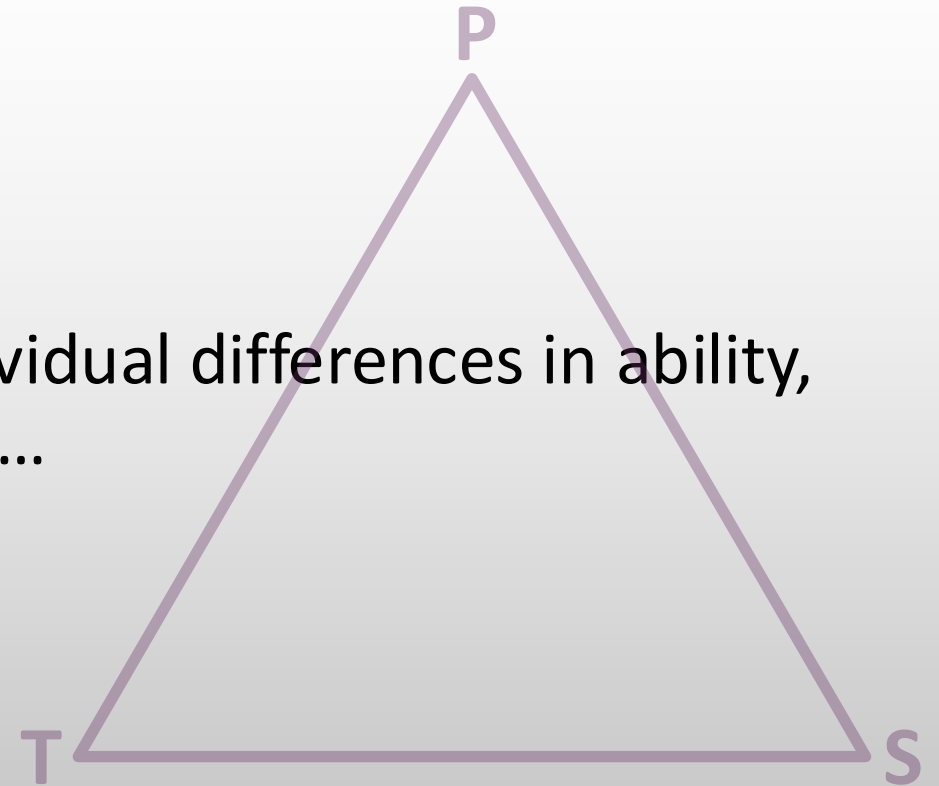
- ... psychological attributes related to individual differences in ability, knowledge, skill, experience, motivation ...

Task

- ... what a person is instructed to do
 - e.g., solving an analogy item

Situation

- ... circumstances in which a task is to be performed



Where to look for Learning



- (Scientific) Realist view: *learning* (as an ability) exists independently of it being measured or not within a PERSON
- Is there such a thing as **Learning Ability**?
 - ... as an ability construct would need to claim “new territory” (e.g., demarcation from intelligence / cognitive ability)
 - New tests – New construct!?
 - **Construct Inflation:** “The first cardinal principle of psychological progress: Give new names to old things.” (McNemar, 1964)

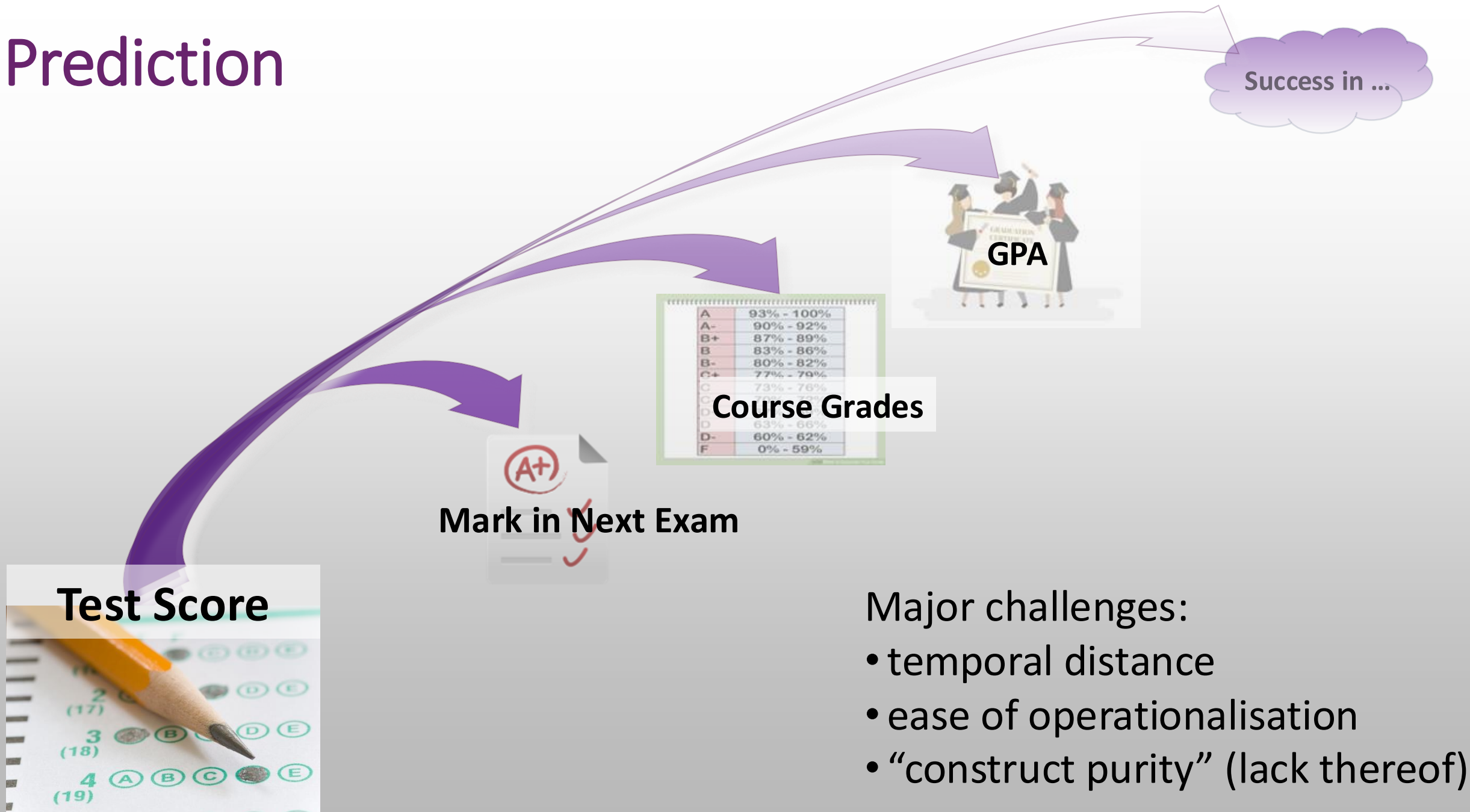
Why are we interested in the assessment of learning?

Decision Making:

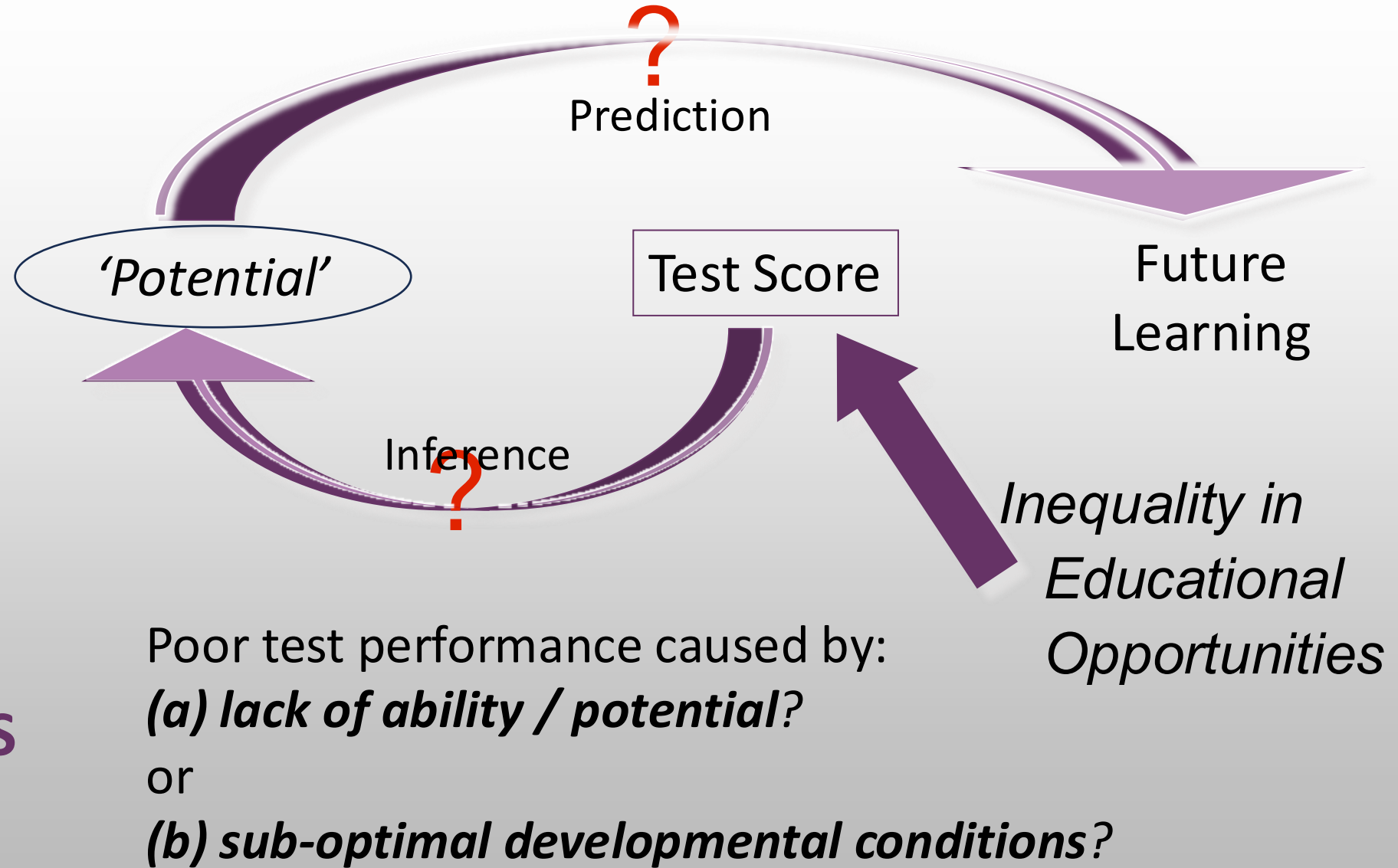
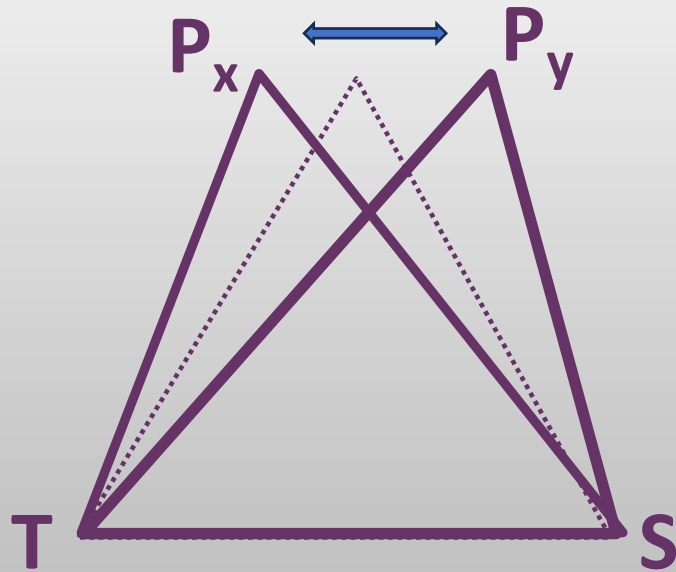
- Selection
- Streaming
- Placement
- ...

based on Prediction

Prediction



Assessment as Between-Person Comparison?



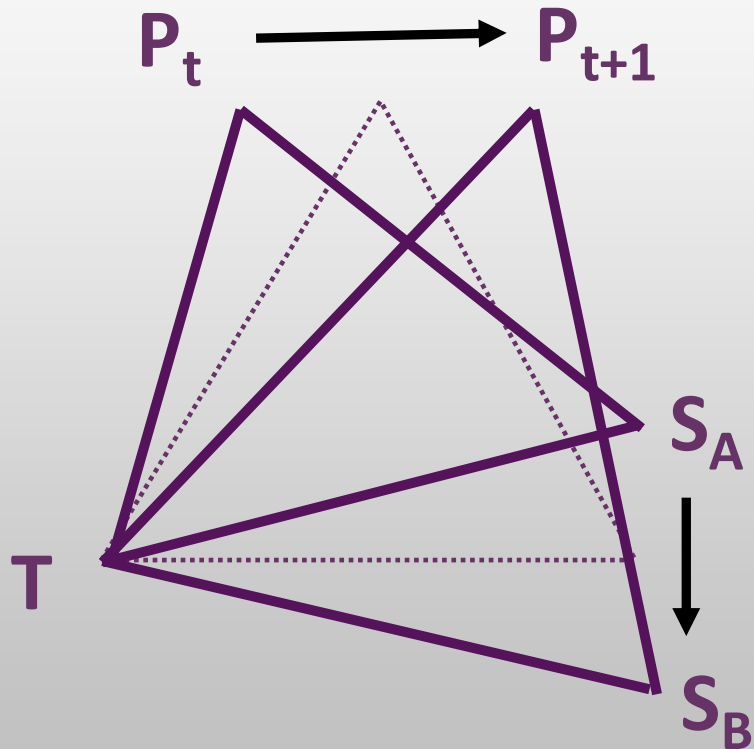
Learning



This Photo by Unknown Author is licensed under CC BY-SA-NC

- ... is a latent, **inferential process** of **intra-individual change** facilitated by **interaction between person** (i.e. the learner), **task** (i.e. what is to be learnt) and **situation** (i.e. circumstances in which learning is expected to take place).
- ... results in a relatively **enduring change** in a person's behaviour or their capability to behave in a purposeful way

Assessment of within-person variability



Learning:

- ... not a person attribute as such,
- ... an emerging quality of purpose-driven and functional behaviour resulting from the interaction between **P**erson attributes, **T**ask demands and **S**ituational characteristics

Learning Ability:

- ... a variable of inter-individual differences in intra-individual variability
 - enables functional adjustment to changing cognitive demands
 - (reasoning) tasks presented under experimentally controlled variation in situational characteristics

What exactly is it that you are claiming to assess?

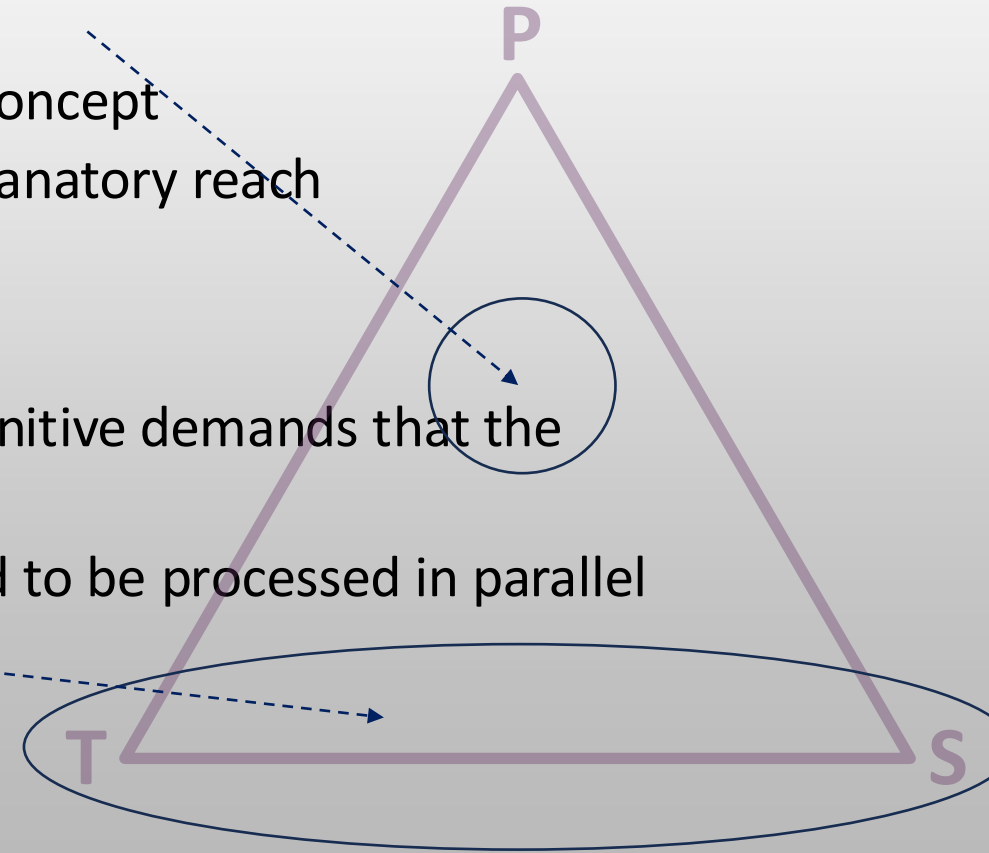
Complexity vs Difficulty

difficulty

- represents the quantifiable level of a person's success in dealing with the complexity imposed by the task and the situational circumstances.
- is experiential, person-bound and by definition, statistical concept
- is a psychometric concept with limited conceptual and explanatory reach

complexity

- is conceptualized as a quality that is determined by the cognitive demands that the characteristics of the task and the situation impose.
- is determined by the number of information cues that need to be processed in parallel when performing the task as instructed
- is a primarily substantive concept



Where to look for Learning

- (Scientific) Realist view: *learning* (as an ability) exists independently of it being measured or not within a PERSON
 - Is there such a thing as **Learning Ability**?
 - ... as an ability construct would need to claim “new territory” (e.g., demarcation from intelligence / cognitive ability)
 - New tests – New construct!?
 - **Construct Inflation**: “The first cardinal principle of psychological progress: Give new names to old things.” (McNemar, 1964)
 - Side effect: Inflation of ‘validities’
 - ... criterion, predictive, face, ecological, statistical, consequential, psychological, ...
 - “Validation”: chasing stars (larger correlations), data-driven
- ... sound familiar?



This Photo by Unknown Author is licensed under CC BY-NC

Validity: The conventional view

‘Validity is an integrated **evaluative judgement** of the **degree** to which empirical evidence and theoretical rationales support the **adequacy** and **appropriateness** of **inferences** and **actions** based on test scores or other modes of assessment’

(Messick, 1989, p. 13)

According to this definition, it’s **not about the test** or the test scores, but it is about:

- the **interpretation** of test scores,
- the **decisions** made based on test scores, and
- the **consequences** of these decisions.



Borsboom & colleagues

validity is...

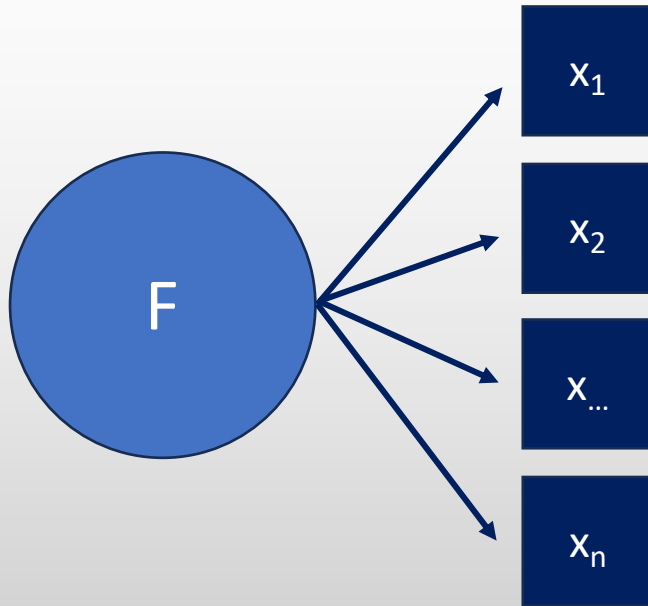
- a property, not a judgment
- a property of instruments, not of inferences
- a function of truth, not of evidence
- the object of validation research, not its result

“A test is valid for measuring an attribute if

- (a) the attribute exists independently of the attempt to measure it, and
- (b) variations in the attribute causally produce variation in the measurement outcomes” (p. 1061).

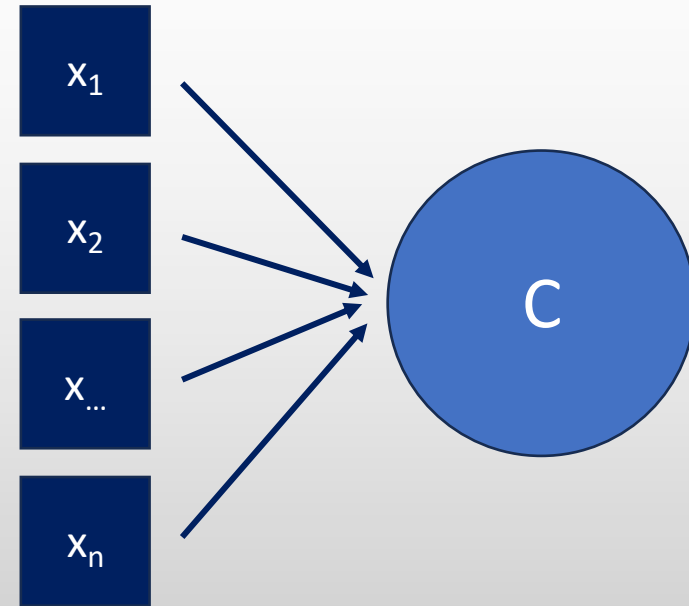


Measurement: What are we getting?



Reflective Measurement Model

- **Factor** as common cause
- Indifference of indicators



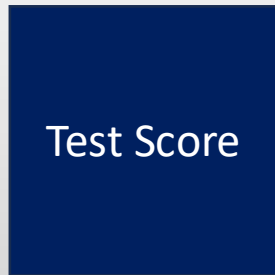
Formative Measurement Model

- **Composite** as shared effect
- Identity of C depends on composition of the set of indicators

Valid or useful?

... it depends

Reflection of a
latent construct?



Formed index of
shared variance?

Validity claims possible, but ...



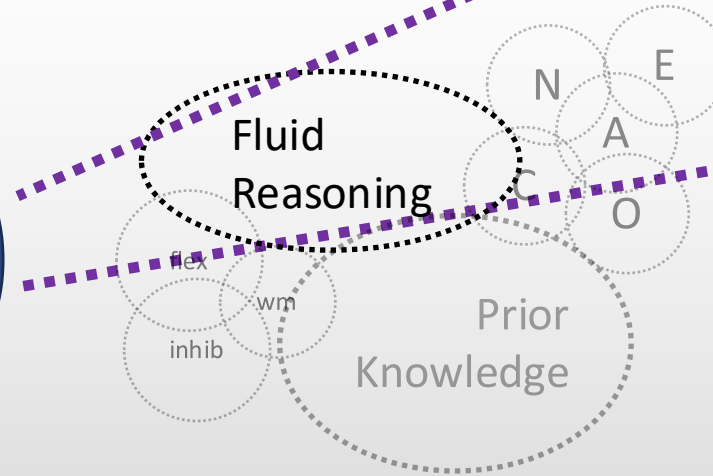
Predictive utility claims possible

BTW: Statements such as, “Test Score **predicts** criterion” are misleading.

use, “Test Score **accounts for** X% of the variance in the criterion” instead.

Construct Validity Lens

- *Limited predictive utility* due to complexity and multi-determinism of “real- life” behaviour

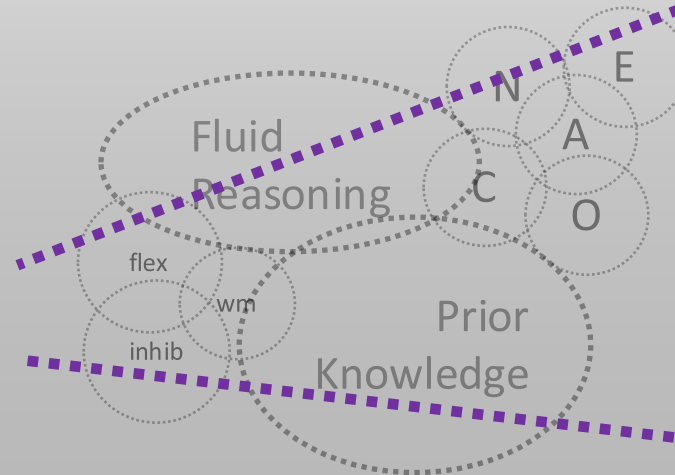


Behaviour (Person A)

Behaviour (Person B)

Predictive Utility Lens

- *Construct impurity* due to complexity and multi-determinism of “real- life” behaviour



Behaviour (Person A)

Behaviour (Person B)

“Predictive Validity”: The blind leading the blind

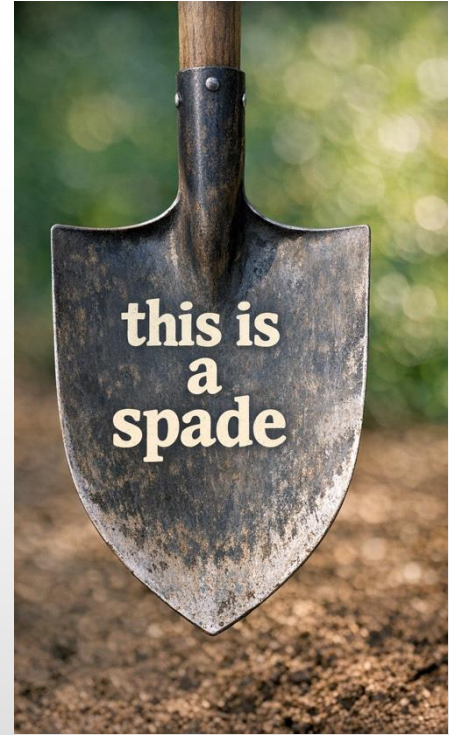
- Predicting whatever one defines as success in future performance is **a-theoretical**
- is based on the **empiricist** principle: ‘past performance being the best predictor of future performance’
- As it is void of a conceptual foundation: any conceptually-based **validity claims are ... invalid.**
- A correlation to a construct-ambiguous criterion cannot be interpreted as indication of construct validity
 - e.g., ‘competency’ is construct-ambiguous. The same competency level can be achieved through primarily fluid reasoning in one person while in another person it might be a matter of prior experiences.



Utility – Validity



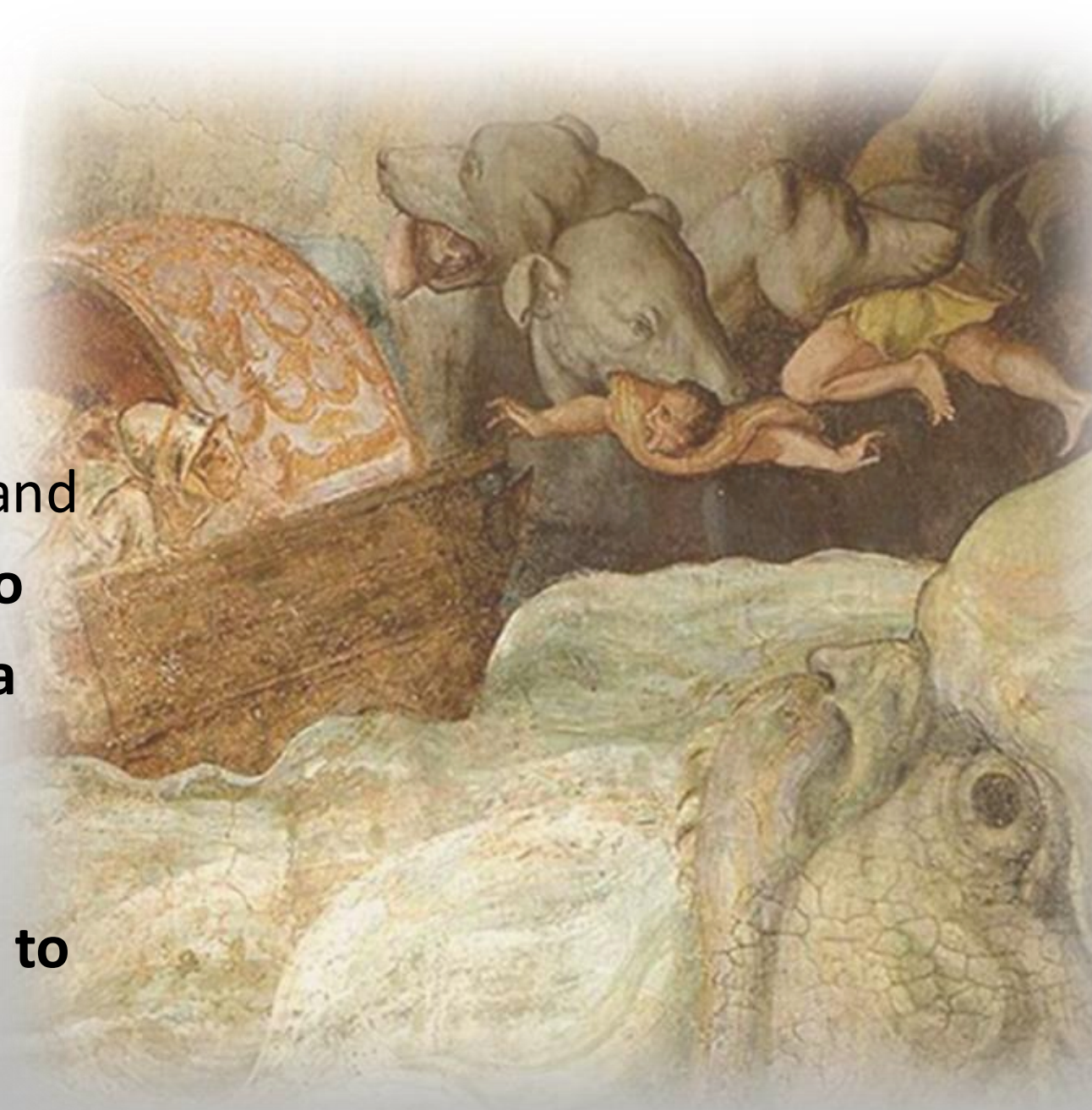
- Scores based on a formative measurement model cannot claim construct validity
- Validity claims need to be based on a reflective measurement model, which in turn requires an *a priory* specification what construct is intended to be measured



- **Predictive Utility** is meaningful; it draws attention to its criterion specificity
- Measurement of dynamic constructs (e.g., Learning Ability or flexibility) requires systematic within-person variation of context demands to affect the interaction between **Person – Task – Situation**
- **Predictive Utility** appears to be much more relevant to practitioners than any of the ambiguous or nebulous validity claims

Validity vs Utility

“Academics are rightly encouraged to disseminate their findings to the widest audience possible, but in doing so often find themselves caught between Scylla and Charybdis: **research that is deemed “too academic” attracts precious little media interests; work that has news value is quickly pounced upon, spun and even abducted to support positions contrary to the researcher’s own.**” (Geoghegan, 2011)





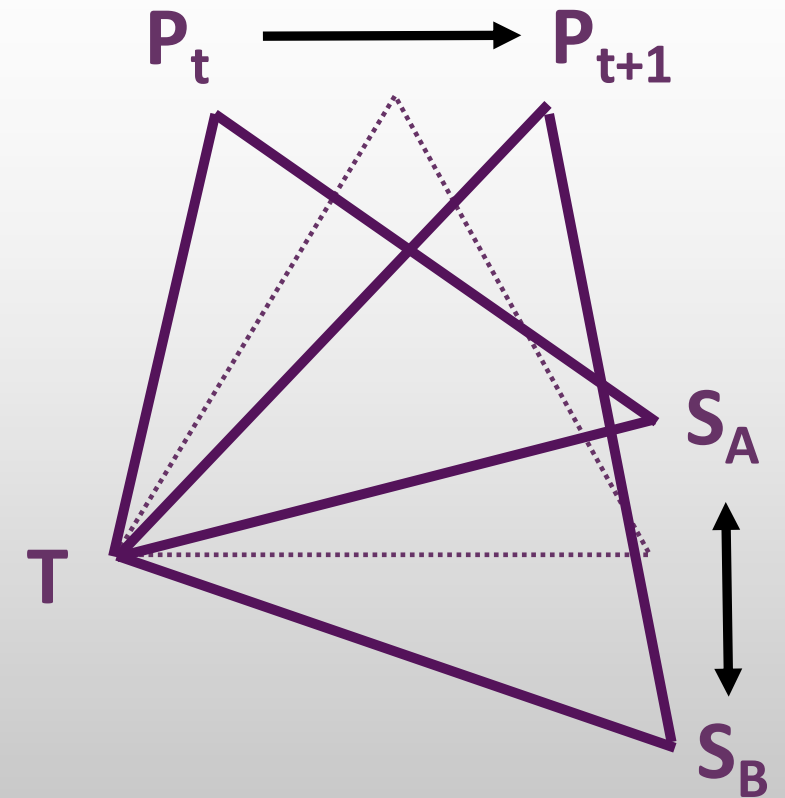
Assessment of Learning via Dynamic Testing

... methodological approach to psychometric assessment that uses systematic variations in

- task characteristics and / or
- situational characteristics

in presenting test items in order to evoke intra-individual variability (in test performance).

Such approach is more appropriate to capture the dynamics of the organisation of human behaviour.



**Inter-individual differences
in
intra-individual variability**

(Beckmann, 2014, p. 310; Guthke & Wiedl, 1996, p. 8; Guthke & Beckmann, 2000, p. 179; Guthke, Beckmann & Wiedl, 2003, p. 225)

Basic assumptions and goals of ...

traditional forms of intelligence testing

- performance = ability = potential
- performance shown independently, i.e. without support is the indicator of ability level
- *Primary goal:* Prediction of future performance based on performance shown in the test (*“future performance best predicted by past performance”*)
⇒ ***developed abilities***

Dynamic Testing

- malleability of performance as indicator of potential
- performance shown in learning phases within the test is indicative of the “true” level of ability as it considers potential
- *Primary goal:* Identifying potential (instead of deficits)
⇒ ***latent capabilities***

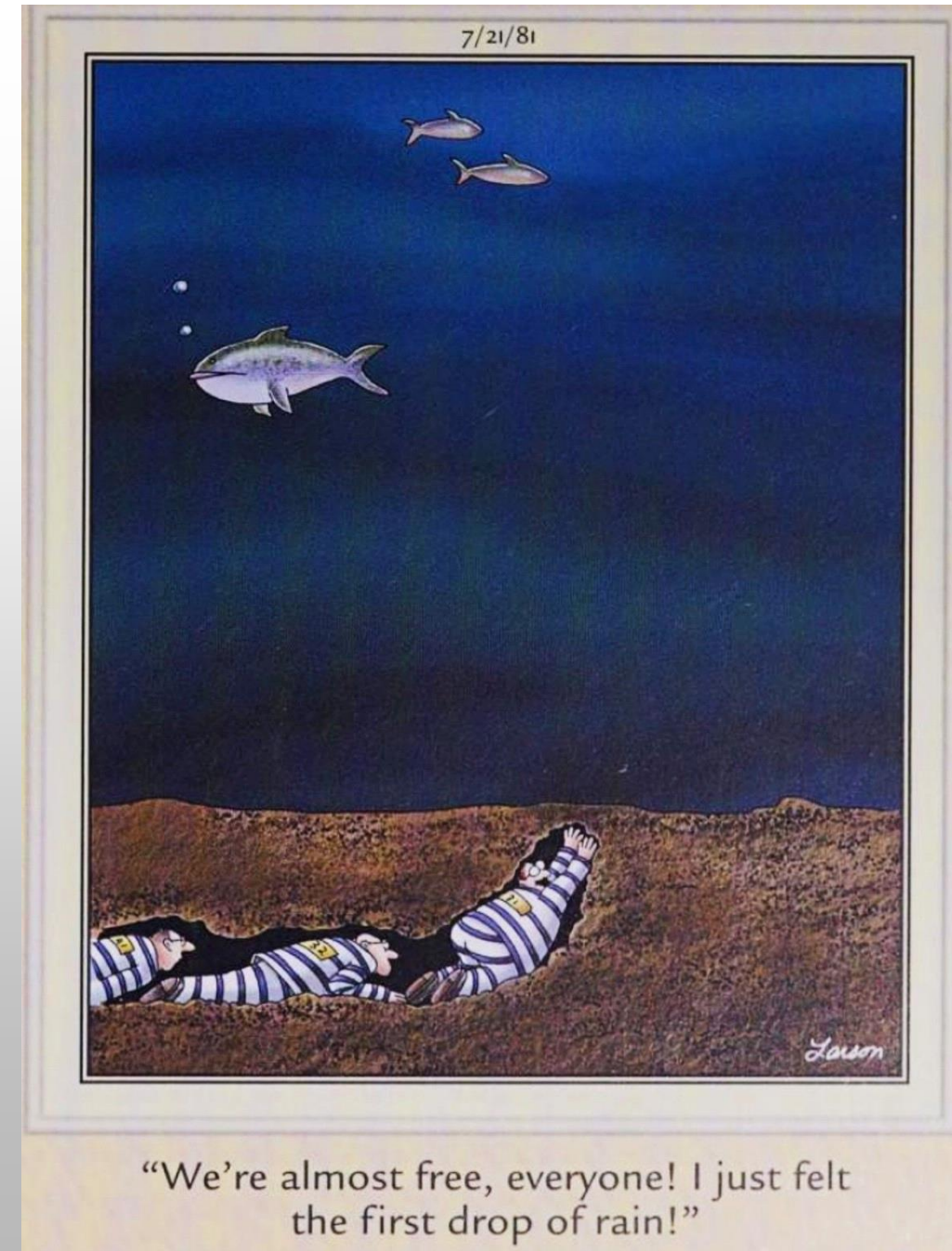
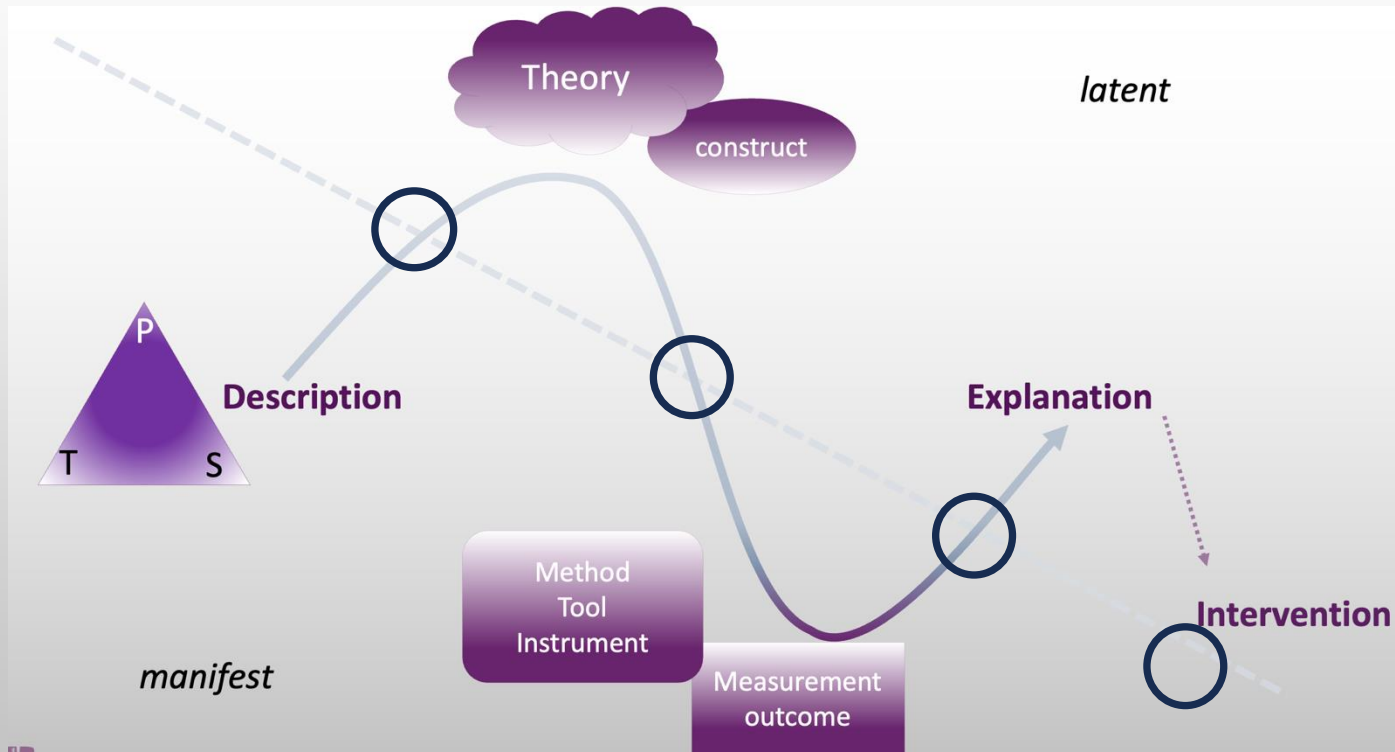
Assessment of learning

- If interested in the **ability to learn** (i.e., to benefit from feedback, cognitive modifiability, intellectual change potential ...),
- the test situation must **provide opportunities to learn** (to benefit from feedback, ...), and
- test scores need to **reflect individual differences** in the extent to which these opportunities were utilised.

Learning & Assessment

- Learning: a latent, **inferential process** of **intra-individual change** facilitated by **interaction between person** (i.e. the learner), **task** (i.e. what is to be learnt) and **situation** (i.e. circumstances in which learning is expected to take place).
- Assessment of learning: capturing inter-individual differences in intra-individual variability
 - watch out for **ergodicity** fallacy!
- Formative vs reflective measurement model
- (Construct) validity $\neq$ predictive utility (and vice versa)
- Description vs explanation
 - mind the distinction between **descriptive causality** and **explanatory causality**
- Complexity vs difficulty

Mind The **Boarder** Crossings



Bibliography – selection (I)

- Beckmann, J. F. (2001). *Zur Validierung des Konstrukts des intellektuellen Veränderungspotentials* [On the validation of the construct of intellectual change potential]. Berlin, Germany: Logos.
- Beckmann, J. F. (2006). Superiority: Always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Psychology*, 23(3), 35–49. <https://doi.org/10.53841/bpsecp.2006.23.3.35>
- Beckmann, J. F. (2014). The Umbrella That Is Too Wide and Yet Too Small: Why Dynamic Testing Has Still Not Delivered on the Promise That Was Never Made. *Journal of Cognitive Education and Psychology*(3), 308–323. <https://doi.org/10.1891/1945-8959.13.3.308>
- Beckmann, J. F., & Dobat, H. (2000). Zur Validierung der Diagnostik intellektueller Lernfähigkeit [The validation of the diagnostic of intellectual learning ability]. *Zeitschrift für Pädagogische Psychologie*, 14, 97–105.
- Beckmann, J. F., & Goode, N. (2017). Missing the Wood for the Wrong Trees: On the Difficulty of Defining the Complexity of Complex Problem Solving Scenarios. *Journal of Intelligence*, 5(2). <https://doi.org/10.3390/jintelligence5020015>
- Beckmann, J. F., & Guthke, J. (1999). *Psychodiagnostik des schlußfolgernden Denkens* [The assessment of reasoning ability]. Göttingen, Germany: Hogrefe.
- Beckmann, J. F., Birney, D. P., & Goode, N. (2017). Beyond Psychometrics: The Difference between Difficult Problem Solving and Complex Problem Solving. *Frontiers in Psychology*, 8, 1739. <https://doi.org/10.3389/fpsyg.2017.01739>
- Beckmann, J. F., Birney, D. P., & Sternberg, R. J. (2024). A Novel Approach to Measuring an Old Construct: Aligning the Conceptualisation and Operationalisation of Cognitive Flexibility. *Journal of Intelligence*, 12(6), 61. <https://doi.org/10.3390/jintelligence12060061>
- Birney, D. P., & Beckmann, J. F. (2022). Intelligence IS Cognitive Flexibility: Why Multilevel Models of Within-Individual Processes Are Needed to Realise This. *Journal of Intelligence*, 10(3), Article 49. <https://doi.org/10.3390/jintelligence10030049>
- Birney, D. P., Beckmann, J. F., & Beckmann, N. (2020). What does within-individual variability in performance trajectories tell us about intelligence? *Personality and Individual Differences*, 157, 109684. <https://doi.org/10.1016/j.paid.2019.109684>

Bibliography – selection (II)

- Birney, D. P., Beckmann, J. F., & Seah, Y.-Z. (2016). More than the eye of the beholder: The interplay of person, task, and situation factors in evaluative judgements of creativity. *Learning and Individual Differences*, 51, 400–408. <https://doi.org/10.1016/j.lindif.2015.07.007>
- Büchli, A., Beckmann, J. F., & Troche, S. J. (2026). Differential Effects of Social and Digital Feedback Interventions in the Figural Analogies Learning Test (FALT). *Journal of Intelligence*, 14(7), 134. <https://www.mdpi.com/2079-3200/14/7/134>
- Du, F., & Beckmann, J. F. (2024). Dynamic testing of language learning aptitude: an exploratory proof of concept study. *Journal of China Computer-Assisted Language Learning*, 4(1), 168–197. <https://doi.org/10.1515/jccall-2023-0028>
- Elliott, J. G., Resing, W. C. M., & Beckmann, J. F. (2018). Dynamic assessment: a case of unfulfilled potential? *Educational Review*, 70(1), 7–17. <https://doi.org/10.1080/00131911.2018.1396806>
- Elliott, J., Stankov, L., Lee, J., & Beckmann, J. F. (2019). What did PISA and TIMSS ever do for us?: the potential of large-scale datasets for understanding and improving educational practice. *Comparative Education*, 55(1), 133–155. [/https://doi.org/10.1080/03050068.2018.1545386](https://doi.org/10.1080/03050068.2018.1545386)
- Guthke, J. (1982). The learning test concept—An alternative to the traditional static intelligence test. *The German Journal of Psychology*, 6, 306–324.
- Guthke, J., & Beckmann, J. F. (2000). *Learning test concept and dynamic assessment*. In A. Kozulin & B. Y. Rand (Eds.), *Experience of mediated learning: An impact of Feuerstein's theory in education and psychology* (pp. 175–190). Oxford, United Kingdom: Elsevier Science.
- Guthke, J., & Beckmann, J. F. (2000). The learning test concept: Application in practice. In C. S. Lidz & J. Elliott (Eds.), *Dynamic Assessment: Prevailing models and applications* (pp. 17–69). Oxford, UK: Elsevier.
- Guthke, J., & Beckmann, J. F. (2001). *Intelligenz als "Lernfähigkeit"—Lerntests als Alternative zum herkömmlichen Intelligenztest* [Intelligence as the ability to learn—Learning tests as alternative to traditional intelligence measures]. In E. Stern & J. Guthke (Eds.), *Perspektiven der Intelligenzforschung. Ein Lehrbuch für Fortgeschrittene* (pp. 137–161). Lengerich, Germany: Pabst.

Bibliography – selection (III)

- Guthke, J., & Beckmann, J. F. (2003). *Dynamic assessment with diagnostic programs*. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 227–242). Washington, DC: American Psychological Association.
- Guthke, J., & Wiedl, K. H. (1996). *Dynamisches Testen. Zur Psychodiagnostik der intraindividuellen Variabilität* [Dynamic testing. On the psycho-diagnosis of intraindividual variability]. Göttingen, Germany: Hogrefe.
- Guthke, J., Beckmann, J. F., & Dobat, H. (1997). Dynamic testing—problems, uses, trends and evidence of validity. *Educational and Child Psychology*, 14, 17–32.
- Guthke, J., Beckmann, J. F., & Stein, H. (1995). Recent research evidence on the validity of learning tests. In J. S. Carlson (Ed.), *Advances in cognition and educational practice. European contributions to the dynamic assessment* (Vol. 3, pp. 117–143). Greenwich, CT: JAI Press.
- Guthke, J., Beckmann, J. F., & Wiedl, K. H. (2003). Dynamik im Dynamischen Testen [Dynamics in dynamic testing]. *Psychologische Rundschau*, 54, 225–232.
- Guthke, J., Beckmann, J. F., Stein, H., Vahle, H., & Rittner, S. (1995). *Adaptive Computergestützte Intelligenz-Lerntestbatterie (ACIL)* [The Adaptive, Computer-Based Learning Test Battery]. Mödlingen: Schuhfried.
- Herrmann, W., Beckmann, J. F., & Kretzschmar, A. (2023). The role of learning in complex problem solving using MicroDYN. *Intelligence*, 100. <https://doi.org/10.1016/j.intell.2023.101773>
- Lidz, C. S., & Elliott, J. G. (Eds.). (2000). *Dynamic assessment: Prevailing models and applications*. Oxford, United Kingdom: Elsevier.
- Tromp, C., & Sternberg, R. J. (2022). Dynamic Creativity: A Person × Task × Situation Interaction Framework. *The Journal of Creative Behavior*, 56(4), 553–565. <https://doi.org/10.1002/jocb.551>