



Key steps in assessment thinking

- **The notion of compromise**
- The notion of programs of assessment
 - Combined purposeful optimization
 - A model
 - Bias reduction
 - An example
- Conclusion



Characteristics of instruments

- Reliability (R)
- Validity (V)
- Educational impact (E)
- Acceptability (A)
- Cost (C)



Utility function

$$U = w_r R \times w_v V \times w_e E \times w_a A \times w_c C$$

- U = Utility
- R = Reliability
- V = Validity
- E = Educational impact
- A = Acceptability
- C = Cost
- W = Weight



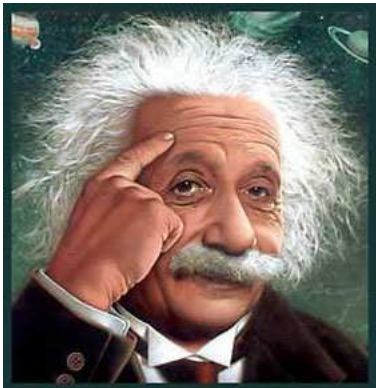
Where do we compromise?

- **Everywhere!**

Validity

- Standardized summative testing is the dominant assessment practice at lower levels of Miller's pyramid

*Not everything that counts can be measured.
Not everything that can be measured counts.*



Albert Einstein



The assessment program

- An arrangement of a purposeful collection of assessment moments, i.e. assessment data points
- The sum is more than the whole
- Should be aligned to the curriculum objectives
- Should foster learning and decision making
- Should combine purposeful compromises.

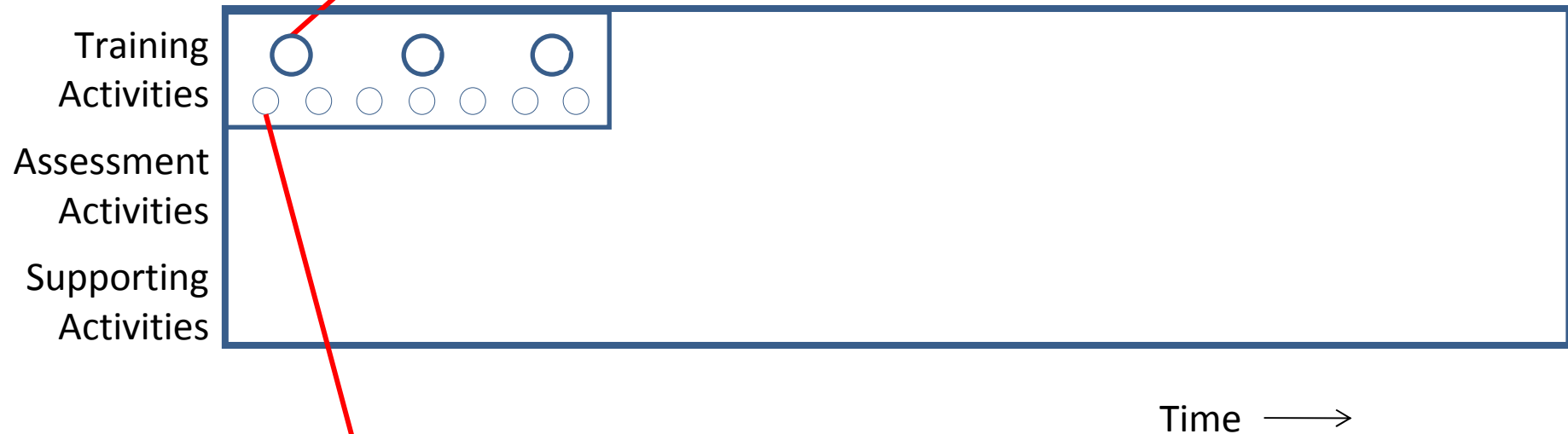


A model of programmatic assessment

- Assumptions:
 - Every individual method is but one (flawed) data point
 - Stakes replaces the formative/summative distinction and is a continuum
 - N of data points should be proportional to stakes
 - Learning model is student-centered, longitudinal, holistic and self-directed.

Artifacts of learning

- Outcome artifacts: Products of learning tasks
- Process artifacts: Learning or working activities

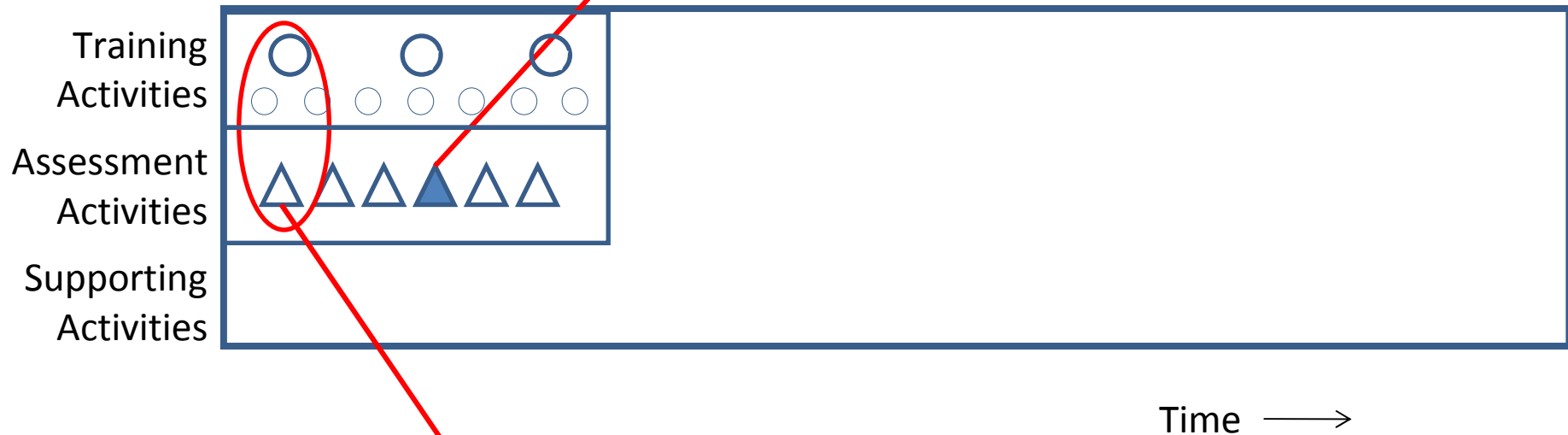


Learning task

- PBL case
- Patient encounter
- Operation
- Project
- Lecture
- Self-study

Certification of mastery-oriented learning tasks

- Resuscitation
- Normal delivery of infant

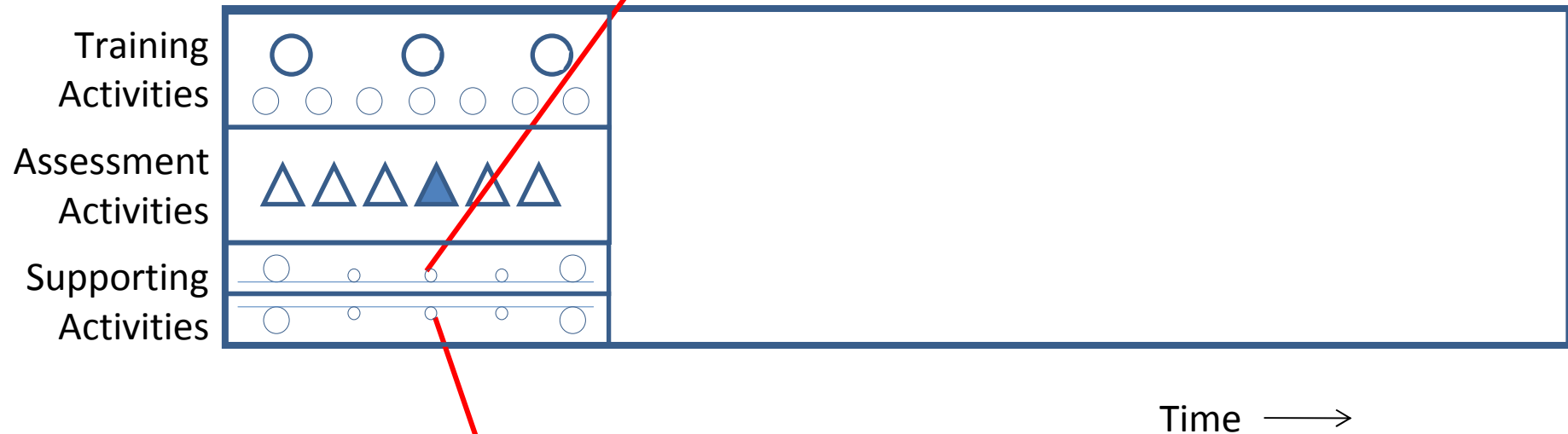


Individual data points of assessment

- Fit for purpose
- Multiple/all levels of Miller
- Learning oriented, Information rich documentation, meaningful (quantitative, qualitative)
- Low stake

(P) Reflective activity by learner

- Interpretation of feedback
- Planning new learning objectives and tasks

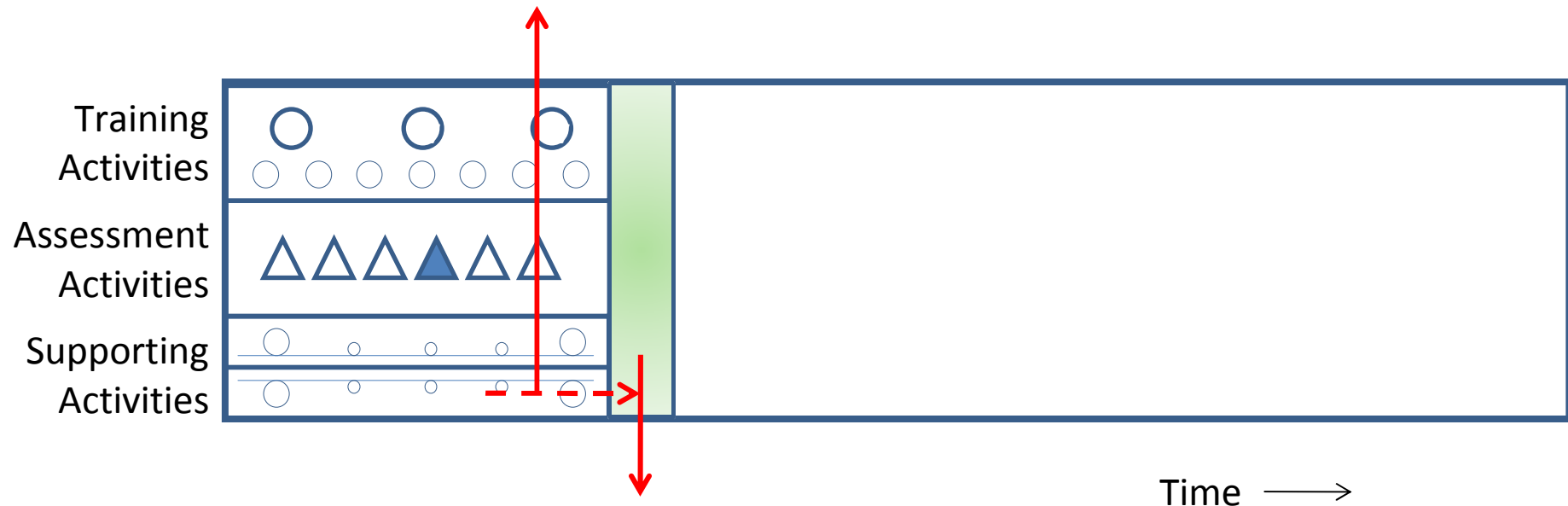


Supportive social interaction

- Coaching/mentoring/supervision
- Peer interaction (interviewing)

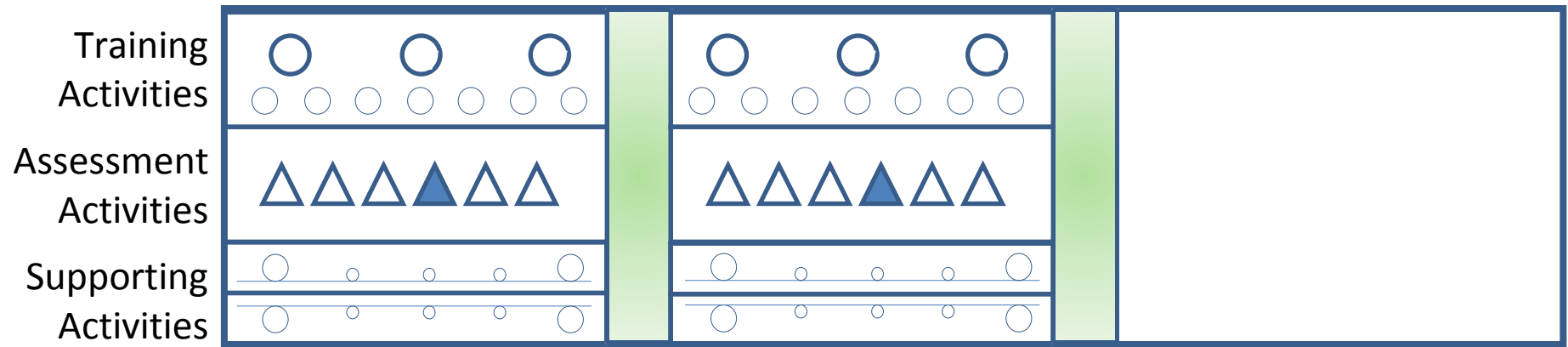
Firewall dilemma

- Dilemma between access to rich information and compromising relationship supporting person(s) and learner

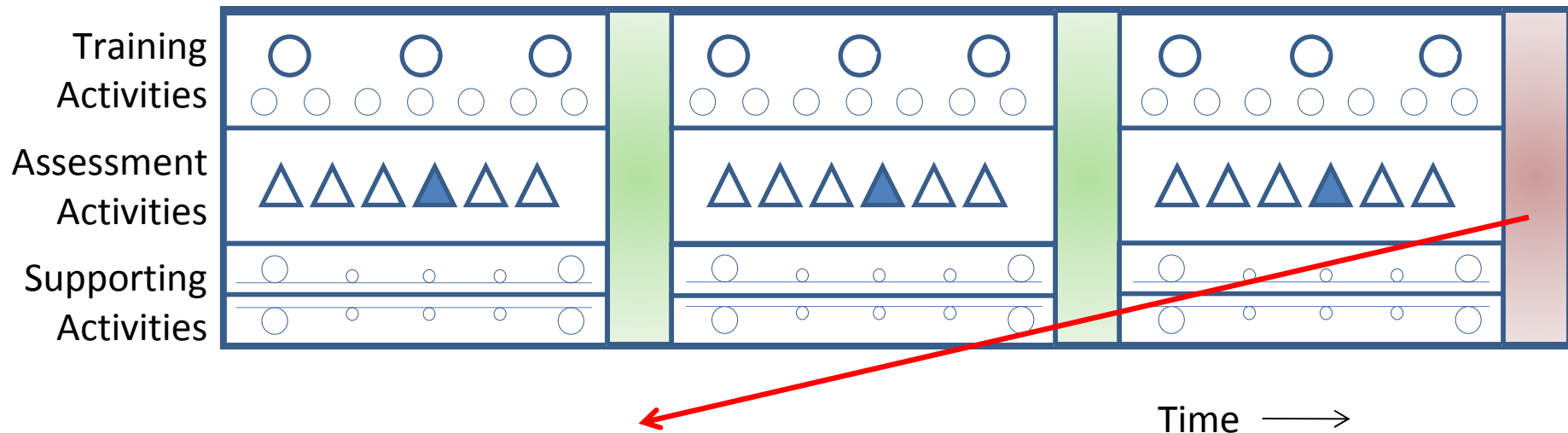


Intermediate evaluation

- Aggregate information held against performance standard
- Committee of examiners
- Decision making: diagnostic, therapeutic, prognostic
- Remediation oriented, not repetition oriented
- Informative
- Longitudinal
- Intermediate stake

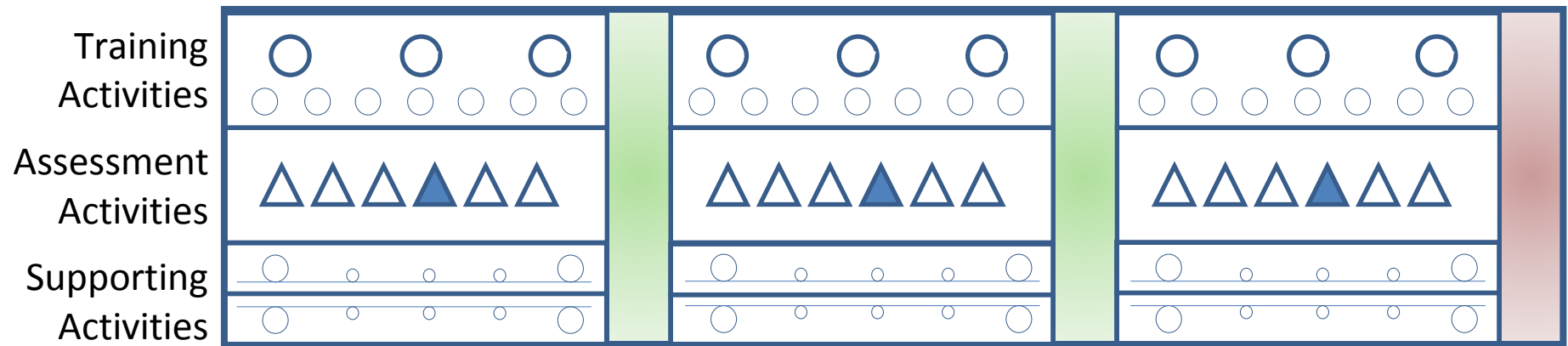


Time →

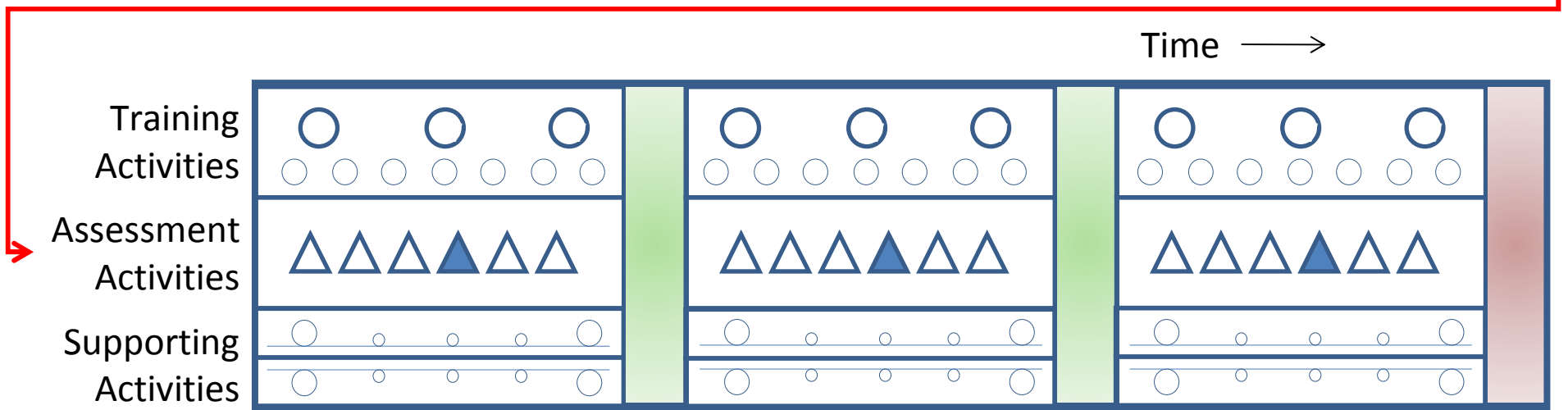
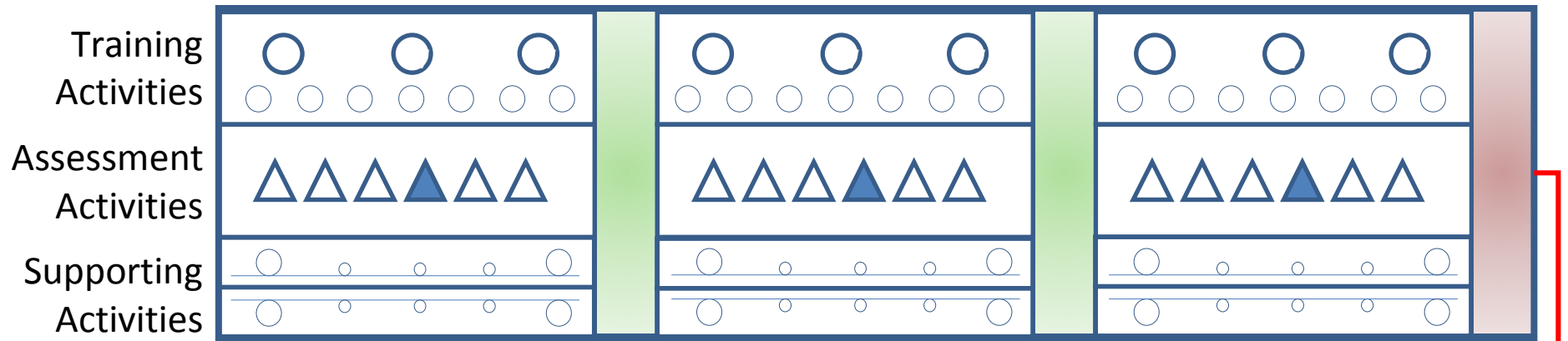


Final evaluation

- Aggregate information held against performance standard
- Committee of examiners
- Pass/fail(/distinction) high stake decision
- Based on many data points and rich information
- Decision trustworthiness optimized though procedural measures, inspired qualitative methodology strategies
- High stake



Time →





How to deal with subjectivity?

- Expert judgment is imperative for assessing complex competencies and when diverse information is to be combined
- Sampling strategies can reduce random error
- Procedural strategies can reduce bias

Strategy to establish trustworthiness	Criteria	Potential Assessment Strategy (sample)
Credibility	Prolonged engagement	Training of examiners
	Triangulation	Tailored volume of expert judgment based on certainty of information
	Peer examination	Benchmarking examiners
	Member checking	Incorporate learner view
	Structural coherence	Scrutiny of committee inconsistencies
Transferability	Time sampling	Judgment based on broad sample of data points
	Thick description	Justify decisions
Dependability	Stepwise replication	Use multiple assessors who have credibility
Confirmability	Audit	Give learners the possibility to appeal to the assessment decision

