Predicting Longer-Term Retention from Learners' Retrieval Practice Performance and its application in education

Maarten van der Velde & Hedderik van Rijn
Multi-session retrieval practice

Learners typically do spaced retrieval practice over multiple sessions. Intervals range from seconds/minutes (within session) to hours/days/weeks (between sessions).

Data set: 25,843 fact learning sequences from university students in a Cognitive Psychology course.
Can we predict longer-term retention from retrieval practice performance?

In terms of ACT-R: Can a single model describe cognition on multiple different timescales (seconds/minutes – hours/days/weeks)?

Are these predictions sufficiently robust to be used in practical applications?
Activation in ACT-R’s declarative memory

Recall probability:
\[ p = \frac{1}{1 + e^{-(A-\tau)/s}} \]

Activation:
\[ A = \ln \sum_{j} t_i^d \]
Activation in ACT-R's declarative memory

Recall probability:

\[ p = \frac{1}{1 + e^{-(A - \tau)/s}} \]

Activation:

\[ A = \ln \sum_j t_j^{-d} \]
Activation in ACT-R’s declarative memory

Recall probability:

\[ p = \frac{1}{{1 + e^{-\frac{(A-1)}{\theta}}}} \]

- 95% for high decay
- 80% for low decay
- 50% retrieval threshold (\( \tau \))
- 5% for high decay

Activation:

\[ A = \ln \left( \sum \frac{t}{i}^{-d} \right) \]
**Activation in ACT-R’s declarative memory**

Recall probability:

\[ p = \frac{1}{1 + e^{-(A-\tau)/s}} \]

- **Low decay**
  - 95%
  - Retrieval threshold \( \tau \)
- **High decay**
  - 20%

Activation:

\[ A = \ln \sum t_j^d \]
Activation in ACT-R’s declarative memory

The model either shows almost no forgetting on short timescales...

Recall probability:
\[ p = \frac{1}{1 + e^{-(A-\theta)/\tau}} \]

- 95% at 0h
- 80% at 2h
- 50% at 4h
- 5% at 24h

Activation:
\[ A = \ln \sum_j t_j^{-d} \]

- Low decay
- High decay

... or almost complete forgetting on long timescales
How does ACT-R do on our multi-session retrieval practice data?
There’s less decay between sessions than expected based on elapsed clock time. 
*Elliott & Anderson (1995); Anderson, Fincham, & Douglass (1999); Pavlik & Anderson (2003)*

Is forgetting slower because of fewer intervening events? 
*See also: context drift in SAM (Mensink & Raaijmakers, 1988), MCM (Mozer et al., 2009)*

“Slowed-clock” model: scale between-session intervals by a factor $h$ (between 0 and 1). Different studies find different values: 0.00046, 0.0172, 0.025, 0.031
*Pavlik, Bolster, Wu, Koedinger, & Macwhinney (2008); Pavlik & Anderson (2008); Pavlik & Anderson (2005); Pavlik & Anderson (2003)*

What is the right value of $h$? Does it depend on the interval?
**Finding a time-variant \( h(t) \) in three steps**

- Bin learning sequences based on between-session interval (here: 20 bins)
- Find best-fitting \( h \) for each bin
- Fit function \( h(t) \)
But there are (at least) two other solutions!

Scaling between-session intervals by a **scaling factor** $h(t)$
- Less interference → shorter “psychological time”.
- Bridge to context-drift-based accounts of forgetting.

**Time-variant decay** $d(t)$
- Lower decay over time → ever stronger “persistence” consolidation.
- See Ribot’s gradient: older memories are more resistant to disruption.

**Time-variant retrieval threshold** $\tau(t)$
- Lower threshold over time → items retrievable at lower activation.
- Suggests an increase in potential invested retrieval effort.
All three parameters change predictably with interval

\[ A = \ln \sum \left( t_{m} + h \times t_{b} \right)^{-d} \]

h(t)
- Activation (A)
- Retrieval threshold (τ)
- Between-session intervals are scaled by a factor h.

d(t)
- Activation (A)
- Retrieval threshold (τ)
- The rate of decay slows over time.

τ(t)
- Activation (A)
- The retrieval threshold drops over time.
**Conflicting evidence in RT**

\[
RT_f = F \cdot e^{-\tau} + t_{er}
\]

\[
RT_c = F \cdot e^{-A_c} + t_{er}
\]
Can predictions from practice replace traditional knowledge tests?

The average Dutch school student receives 102 grades per school year

De Correspondent, 2023
Can predictions from practice replace traditional knowledge tests?
Model-based assessment of mastery

Define mastery in terms of projected retention, using the same threshold-based memory model:
Memory activation predicts test performance

Memory activation at end of study session
Model-based predictions outperform “raw” measures.
Which of the following tissues is a secondary lymphoid organ?

1. Thymus
2. Spleen
3. Bone marrow
4. Liver
Cognitive Psychology

- Students have to know all glossary items by heart (of some of these, the definition will be given on the exam, with the answer being the term)
  - Account for 30% of total grade

- Previous years: students could use MemoryLab to study (=> 6.2/10 grade)

- 2023-2024: if two Mastery Credits per chapter were obtained, students were guaranteed a 7.5/10 grade on the fact-part of the exam.
  - The exam could also be taken without MemoryLab studying.
Students Start Earlier!

Cumulative usage

Date

Cumulative responses

Data: RuG Cognitive Psychology 2023/2024
Prediction accuracy increases

Data: RuG cognitive psychology 2023/2024
Students Distribute Learning
Proof of the Pudding

Grade: 6.2/10  
Passing rate: 60%

Grade: 7.4/10  
Passing rate: 83%

Students gave very positive reviews!
Predicting Longer-Term Retention from Learners' Retrieval Practice Performance and its application in education

Maarten van der Velde & Hedderik van Rijn