AIAI-EVD-004 | CHAPTER 04 | Research notes

What does a high score actually test?

A closer look at the unit of evaluation: new records can still belong to familiar groups.

Jessica Fogarty | AIAI.science
Public editorial edition 1.0 | 27 October 2026

Take away three things

1. Define what must be genuinely unseen.

2. Inspect relationships across the evaluation boundary.

3. Report the conditions before generalising the result.

The hidden unit of independence

A dataset can contain several records from one person, device, organisation or experiment. Separating rows does not necessarily separate their originating groups.

If the intended deployment involves entirely new groups, the evaluation needs to address that target. Otherwise a model may benefit from relationships that will not be available in the intended use.

What the exploratory work examined

The reviewed draft compares record-level and group-disjoint evaluation in a deliberately constructed nearest-neighbour example. It keeps the comparison transparent so the effect of group overlap can be inspected.

The construction illustrates an established evaluation failure mode. Its measurements belong to the stipulated synthetic setting; they are not a real-world accuracy estimate or a test of the full architecture.

Questions for an evaluation brief

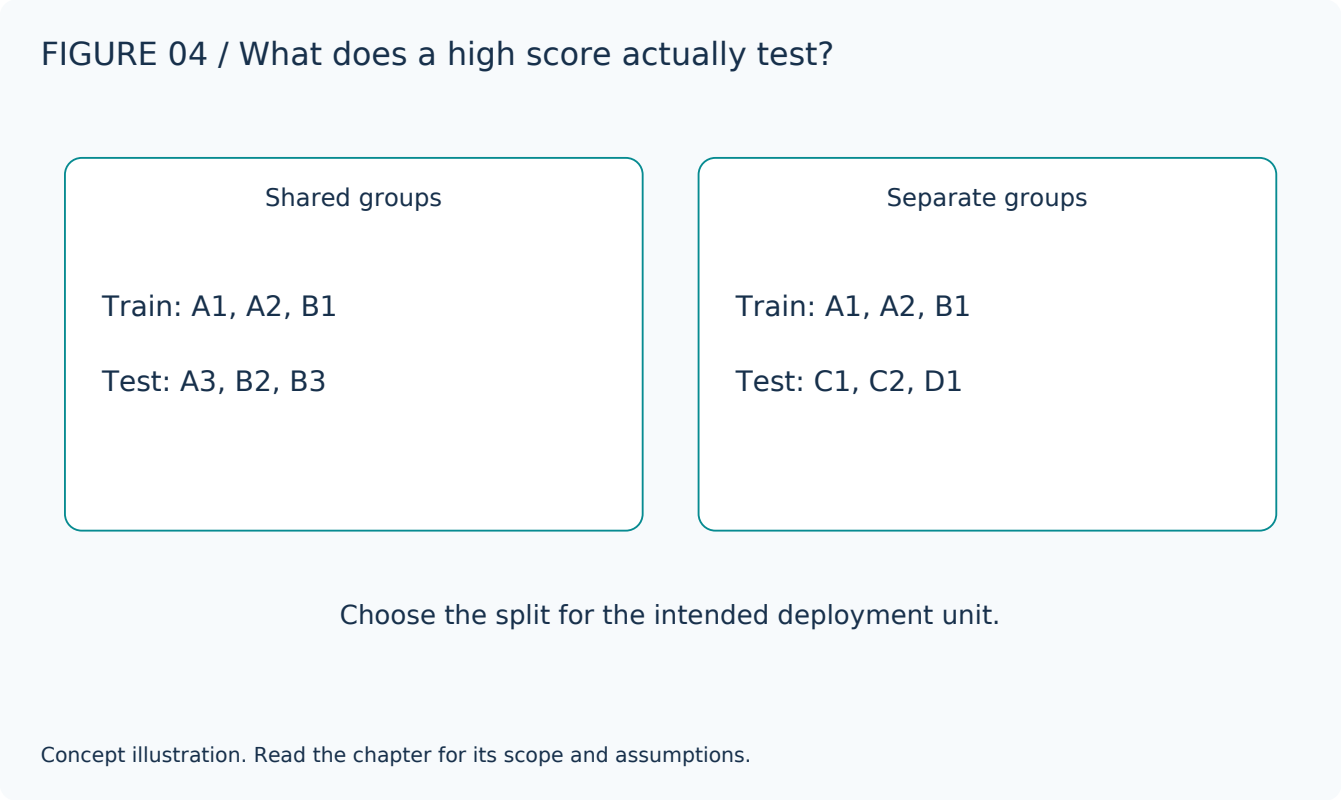
State the intended deployment population, the grouping rule, what information is available at prediction time and how overlap was checked. Keep validation data outside the decisions used to fit or select the system.

When performance changes, investigate the target and conditions before treating the difference as proof of universal capability or failure.

The portfolio connection

This question informs ProvenResearch, the UMT-Ω evaluation direction and the way AIAI.science describes prototype evidence. A result becomes more useful when its scope can be inspected.

Figure 04. The idea in one view



Conceptual explanation with illustrative inputs where shown. This figure is not an empirical result or a validation of a deployed system.

Basis and scope

Public explanatory note informed by exploratory UMT-Ω research drafts reviewed 27 October 2026. The underlying studies use synthetic constructions, are not peer reviewed or externally replicated, and do not evaluate the complete OMNI CORTEX architecture. No private draft or dataset is included here.

Read online and explore related work

Open this chapter online

The online edition includes interactive representations where useful. The chapter remains complete as a static explanation.