

## Mínimo viable para obtener las estadísticas con un usuario:

Mínimo se necesitan 3 sesiones para comprobar la mejora de retención, cabe recalcar que no se mide sobre el mismo test o resumen, se mide sobre el resultado obtenido sobre los test.

Sesión 1: Toma de contacto y resultado inicial.

Sesión 2: Ya conoces la herramienta y obtienes tu primera comparación de entre tu resultado pasado y el actual.

Sesión 3: Ya veras tu trayectoria y mejora conforme aplicas la técnica.

## Cómo calculamos su retención:

Definición de la Métrica: El "Índice de Retención" (IR)

Para cada sesión de estudio (lectura + test), obtendremos dos datos:

1. Score-inmediato: Calificación del test inmediato (comprensión).
2. Score-espaciado: Calificación del test espaciado (retención real).

La Retención Individual (IR) de esa sesión se calcula así:

$$IR = \text{Score-espaciado} / \text{Score-inmediato}$$

*(Esto te dice qué porcentaje de lo que el usuario aprendió inmediatamente fue capaz de conservar tras 48 horas).*

## Cálculo de la Mejora Individual (Delta Δ)

Para medir la mejora, el usuario debe realizar al menos 2 sesiones (preferiblemente 3).

- Sesión 1 (S1): Es su línea base (sin apoyo profundo de IA o primer contacto).
- Sesión 2+ (S2): Donde ya aplica los resúmenes de IA y el método espaciado.

Fórmula de Mejora Individual:

$$\text{Mejora} = IRS_2 - IRS_1$$

Ejemplo real:

- En la Sesión 1: El usuario sacó 90 en el inmediato y 60 a las 48h. Su IR=0.66 (retuvo el 66%).
- En la Sesión 2: El usuario sacó 90 en el inmediato y 80 a las 48h. Su IR=0.88 (retuvo el 88%).
- Resultado: Su mejora individual es de 22%. ¡Hipótesis validada!

## Fundamento científico:

Están fundamentados en dos pilares de la **Psicología Cognitiva** y la **Ciencia del Aprendizaje**

### El Índice de Retención de Ebbinghaus

Las operaciones se basan en la **Curva del Olvido** de Hermann Ebbinghaus. Él estableció que la memoria no es estática, sino que decae de forma exponencial.

- **Fundamento:** Ebbinghaus utilizó el concepto de "**Ahorros**" (**Savings**), que es básicamente la diferencia entre el tiempo que tardas en aprender algo la primera vez vs. la segunda.
- **Nuestra aplicación:** Al comparar el Score-espaciado con el Score-inmediato, está midiendo matemáticamente el **Índice de Decaimiento**. Si la app reduce ese decaimiento, está "aplanando la curva", un concepto estándar en estudios de neurociencia educativa.

### El "Testing Effect" (Roediger & Karpicke, 2006)

El estudio clave que debes citar es "*Test-Enhanced Learning*". Estos investigadores demostraron que los estudiantes que realizan exámenes de recuperación (como tus quizzes) retienen significativamente más información a largo plazo que aquellos que solo releen (lectura pasiva).

- **La base de tu 20%:** En sus experimentos, los grupos que hicieron "tests" mostraron mejoras de retención de entre el **15% y el 25%** en comparativa con grupos de estudio tradicional tras periodos de 2 días a una semana. De aquí sale el sustento científico de que tu meta del **20%** es realista y alcanzable.

### El Modelo de Evaluación de Aprendizaje de Kirkpatrick

Aunque es un modelo empresarial/educativo, el **Nivel 2 (Aprendizaje)** mide precisamente la diferencia entre "Antes" y "Después".

- **Fundamento:** Establece que para medir el aprendizaje real, se debe realizar una evaluación previa (Pre-test) y una posterior (Post-test).
- **Tu aplicación:** Tu "Sesión 1" actúa como el **Pre-test** (Control) y las sesiones con la IA actúan como el **Post-test**(Experimental). La resta de ambos resultados es la métrica estándar para reportar el "Gain Score" o Ganancia de Aprendizaje.

*"El cálculo del porcentaje de mejora se basa en el Gain Score Analysis, comparando el desempeño del usuario contra su propia línea base. Esta métrica está validada por los estudios de Roediger & Karpicke (2006) sobre el 'Testing Effect', los cuales demuestran que la evaluación activa incrementa la retención a largo plazo en rangos similares al 20% propuesto en nuestra hipótesis."*

## Resumen:

- **¿Es lógico?** Sí, es una regla de tres/porcentaje simple.
- **¿Es científico?** También. Estás usando un **Diseño Intra-sujeto** (el usuario es su propio control), que es el método más riguroso para eliminar variables externas como la inteligencia previa o el grado académico.

## Cálculo global de la mejora:

Dividiremos los 20 días en dos fases para obtener la comparativa:

- **Semana 1 (Línea Base):** El usuario lee un PDF sin herramientas de IA y hace un test 48h después.
- **Semana 2 y 3 (Uso de ReadFlow):** El usuario usa tu app (Resumen + Test Inmediato + Test 48h).

¿Cuántos resúmenes y tests debe hacer cada usuario?

Para que los datos sean estadísticamente válidos en 20 días:

- **Total de sesiones:** 3 sesiones completas por usuario.
- **Cada sesión incluye:** 1 lectura de PDF + 1 Test Inmediato + 1 Test Espaciado (48h).
- **Por qué 3:** La primera sesión suele tener sesgo por "novedad". El promedio de 3 sesiones da un dato real de retención.

Sistema de Calificación (Score) por tipo de Test

Como tenemos tests híbridos, debemos normalizar el puntaje a una escala de **0 a 100 puntos**:

1. **Opción Múltiple (30% del peso):** Calificación automática binaria (Correcto/Incorrecto).
2. **Completar palabras (30% del peso):** Calificación por coincidencia de palabras clave (Keywords).
3. **Redacción libre (40% del peso):** Aquí es donde brilla **Gemini**. Envías el texto del usuario a la API y le pides: *"Compara este resumen del usuario con el texto original y califica del 1 al 40 cuántos conceptos clave retuvo"*.

## Operaciones Matemáticas para medir el % de Mejora

Para probar la hipótesis, usaremos la fórmula de **Retención Diferida**.

A. Cálculo de la Retención Individual (R):

R=Puntaje Test 48h

B. Cálculo de la Mejora ( $\Delta$ ):

Necesitamos comparar la retención **sin** la app (Rcontrol) contra la retención **con** nuestra app (Rapp).

$$\% \text{Mejora} = \left( \frac{\bar{R}_{app} - \bar{R}_{control}}{\bar{R}_{control}} \right) \times 100$$

Donde  $\bar{R}$  es el promedio de los 40 usuarios.

¿Cómo asegurar el 20% de mejora? (lógica de IA)

La clave no está en el test inmediato, sino en el **test de las 48 horas**. Tu base de datos debe registrar:

| Usuario | Test Inmediato (T0) | Test 48h (T48) | Índice de Olvido |
|---------|---------------------|----------------|------------------|
| Sin App | 80 pts              | 50 pts         | -30 pts          |
| Con App | 90 pts              | 75 pts         | -15 pts          |

**El éxito consiste en demostrar que con ReadFlow, el "olvido" es menor.** Si sin la app olvidan 30 puntos y con la app solo olvidan 15, has mejorado la retención significativamente.