



## Héctor Enrique Banquez Buendía

Doctorando en Educación Matemática - UPEL

Magister en Recursos Educativos Digitales – UDC

Docente de Matemáticas y Líder de Plataforma Educativa - Colegio Biffi de Cartagena  
bahecbu@gmail.com

Comunicación de innovación · UED

Universidad de los Andes · 18 de agosto de 2026

# El error como aprendizaje

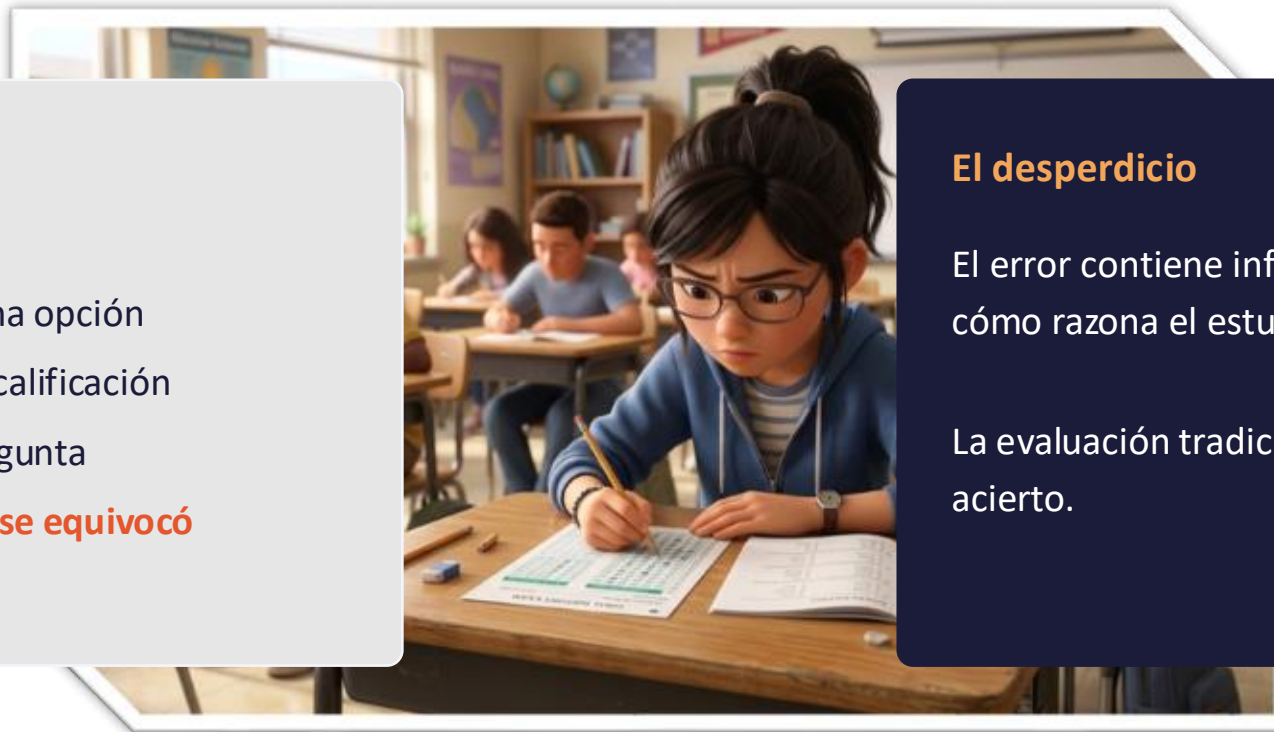
Pensamiento aleatorio y preguntas Saber 11 en entornos digitales



# Una escena conocida en el aula

## Lo que suele pasar

- El estudiante marca una opción
- Ve “incorrecto” y una calificación
- Pasa a la siguiente pregunta
- **Nunca sabe POR QUÉ se equivocó**



## El desperdicio

El error contiene información muy valiosa sobre cómo razona el estudiante.

La evaluación tradicional la borra: solo cuenta el acierto.

*“Los errores pueden ser trampolines de indagación.” — Borasi (1996)*

Y en el pensamiento aleatorio el problema se agrava: las intuiciones sobre el azar son tenaces y sobreviven a la clase.

# ¿Y si cada opción incorrecta fuera un diagnóstico?

1

## Distractor con intención

Cada opción incorrecta se diseña sobre una concepción errónea documentada.

2

## Retroalimentación que explica

Al fallar, el recurso nombra el razonamiento que llevó al error.

3

## Reflexión inmediata

El estudiante escribe qué pensó, en el momento del error, antes de avanzar.



*El distractor deja de ser “la trampa” y pasa a ser una ventana al razonamiento. Y esa ventana no solo sirve para mirar: sirve para intervenir*

Sadler (1998) · Treagust (1988)

# Ítems tipo Saber 11 en un entorno interactivo

15

Ítems de pensamiento aleatorio

3

DBA: muestreo, dispersión, probabilidad

4

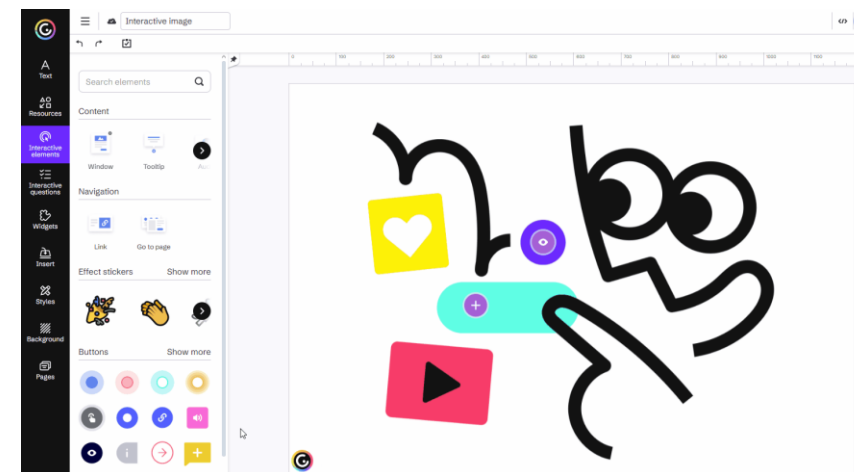
capas por ítem: pregunta, error, teoría, reflexión

1

solo intento: la primera elección cuenta

**Construido en Genially, sin programación.**

- Registra qué opción marcó cada estudiante y exporta los datos
- Captura anónima: sin nombres, solo códigos
- La reflexión escrita funciona como “peaje” antes de avanzar
- Narración en audio y apoyo visual en cada explicación



# El recorrido del error, paso a paso

1

## Marca una opción

Un solo intento: queda registrada.

2

## Recibe el diagnóstico

Se le nombra el razonamiento del error.

3

## Revisa la teoría

Explicación visual y en audio.

4

## Escribe su reflexión

Antes de pasar a la siguiente.

Retroalimentación orientada al proceso:  
Hattie y Timperley (2007) · Shute (2008)

### (a) Ítem con distractores diagnósticos

**Pregunta 2**

En una fábrica operan dos máquinas con tasas de falla muy distintas: la Máquina A (nueva) produce 1.000 tornillos/hora y falla históricamente el 1 %; la Máquina B (antigua) produce 1.000 tornillos/hora y falla el 15 %. El ingeniero de calidad toma 100 tornillos de una caja común de salida, donde se mezclan los productos de ambas máquinas, sin distinguir de cuál vienen.

**Esquema de Muestreo para Control de Calidad**



¿Por qué este método NO es el más adecuado para estimar el total de defectos?

A. Porque la muestra es demasiado pequeña respecto a la producción total.    B. Porque debió tomar los 100 tornillos solo de la Máquina B, que es la que más falla.    C. Porque al no estratificar por máquina, la muestra pierde precisión.    D. Porque el muestreo aleatorio simple siempre es más preciso que cualquier otro método.

### (b) Retroalimentación por distractor (opción correcta revelada)

El punto clave de este ítem no es el número de tornillos, sino cómo se eligen.

- Si marcaste A, caíste en el “mito del tamaño suficiente”: 100 puede bastar matemáticamente; el problema es el método de selección, no la cantidad.
- Si marcaste B, eso sería un muestreo intencional (a propósito hacia la peor máquina), que no representa el total de la fábrica.
- Si marcaste D, cuidado con el “siempre”: en estadística casi nunca aplica; cuando hay subgrupos muy distintos, el aleatorio simple es menos preciso que el estratificado.

Revisar teoría    Consigna de reflexión

estimar el total de defectos?

A. Porque la muestra es demasiado pequeña respecto a la producción total.    B. Porque debió tomar los 100 tornillos solo de la Máquina B, que es la que más falla.    C. Porque al no estratificar por máquina, la muestra pierde precisión.    D. Porque el muestreo aleatorio simple siempre es más preciso que cualquier otro método.

### (c) Capa de teoría multimodal («Revisar teoría»)

El punto clave de este ítem no es el número de tornillos, sino cómo se eligen.

- Si marcaste A, caíste en el “mito del tamaño suficiente”: 100 puede bastar matemáticamente; el problema es el método de selección, no la cantidad.
- Si marcaste B, eso sería un muestreo intencional (a propósito hacia la peor máquina), que no representa el total de la fábrica.
- Si marcaste D, cuidado con el “siempre”: en estadística casi nunca aplica; cuando hay subgrupos muy distintos, el aleatorio simple es menos preciso que el estratificado.

Revisar teoría    Consigna de reflexión

No todas las muestras aleatorias son iguales. Cuando una población está dividida en grupos muy diferentes (estratos), como una fábrica con máquinas nuevas y viejas, usar un muestreo simple puede ser un error. Es mejor usar un muestreo estratificado para asegurar que todos los grupos estén representados proporcionalmente.



### (d) Consigna de reflexión escrita en el momento del error

El punto clave de este ítem no es el número de tornillos, sino cómo se eligen.

- Si marcaste A, caíste en el “mito del tamaño suficiente”: 100 puede bastar matemáticamente; el problema es el método de selección, no la cantidad.
- Si marcaste B, eso sería un muestreo intencional (a propósito hacia la peor máquina), que no representa el total de la fábrica.
- Si marcaste D, cuidado con el “siempre”: en estadística casi nunca aplica; cuando hay subgrupos muy distintos, el aleatorio simple es menos preciso que el estratificado.

Revisar teoría    Consigna de reflexión

Con tus palabras, explica por qué, cuando dos máquinas fallan de forma muy distinta, conviene separar por máquina (estratificar) en lugar de tomar todo de la caja mezclada. Si elegiste otra opción, ¿qué te hizo pensar que era la correcta?

Escribe aquí tu respuesta.

Enviar

Diagnóstico → reparación → articulación → avanzar

# Cada opción incorrecta tiene un nombre

Dos máquinas fallan 1 % y 15 %. El ingeniero toma 100 tornillos de una caja donde se mezclan. ¿Por qué este método no es el más adecuado?



A “La muestra es demasiado pequeña”

*Mito del tamaño suficiente*

B “Debió tomar todo de la máquina B”

*Muestreo intencional*

C “No estratificar hace perder precisión”

*Respuesta correcta*

D “El aleatorio simple siempre es mejor”

*El absoluto “siempre”*

Diseñar así toma más tiempo — pero convierte la prueba en un mapa del razonamiento del grupo.

# Dos ciclos: probar, aprender, rediseñar

## Ciclo 1 • Pilotaje

Un curso • 29 estudiantes

### Qué falló:

- Podían intentar hasta acertar → adivinaban
- La reflexión en papel, al final, llegaba fría



Validación por  
colegas expertos

## Ciclo 2 • Aplicación

Dos cursos • 54 estudiantes

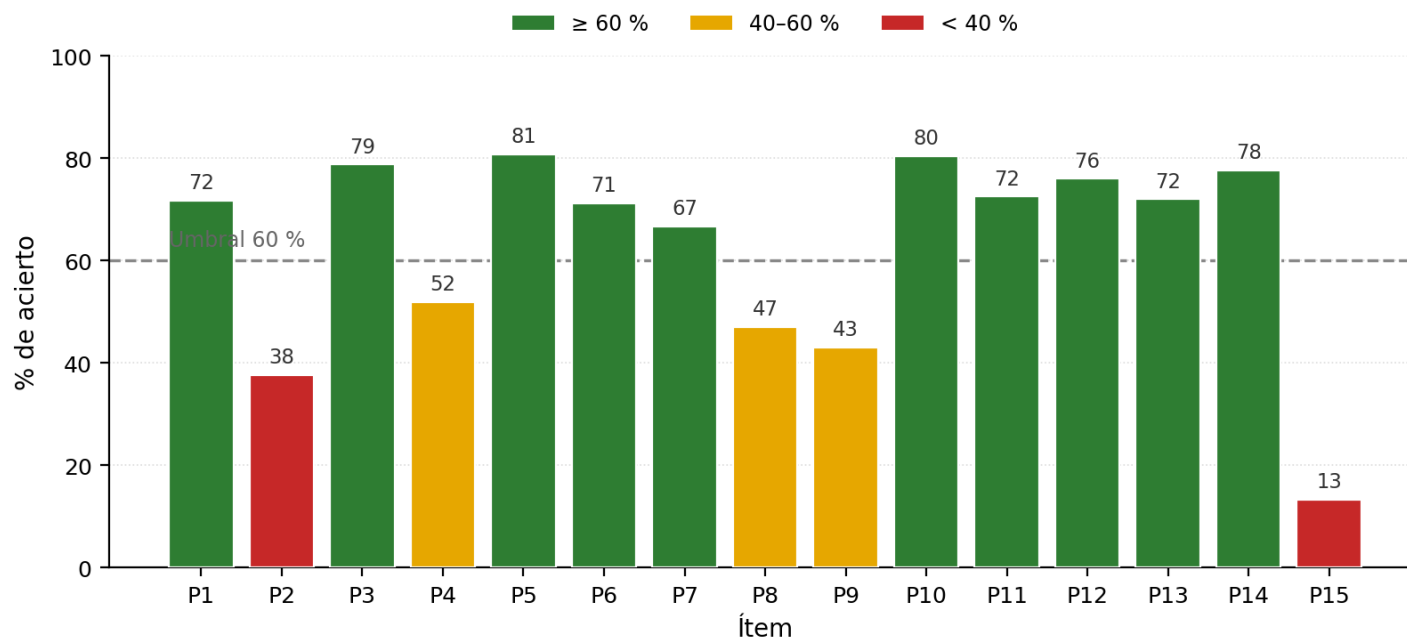
### Qué cambié:

- Un solo intento: la primera elección cuenta
- Ítems corregidos y opciones más precisas
- Reflexión escrita dentro del recurso, al fallar



*El pilotaje no “salió mal”: su función era mostrarme qué corregir. Esa es la lógica del diseño iterativo.*

# La dificultad no estaba donde la esperaba



## El patrón

Los cinco ítems más difíciles no comparten un tema.

Comparten una exigencia:

**coordinar dos parámetros a la vez**

*Media y desviación · tamaño y método · valor y probabilidad*

**Leer un dato es fácil. Coordinar dos es donde se rompe el razonamiento estadístico.**

Promedio del grupo: 8,4 de 15 · Ítem más difícil: esperanza matemática (13 % de acierto)

# Los distractores dibujaron el mapa del razonamiento

45 %

**creyó que el problema era el tamaño de la muestra**  
*y no el método de selección*

44 %

**decidió solo con el promedio**  
*ignorando por completo la dispersión*

57 %

**confundió la pérdida esperada con el costo del seguro**  
*el ítem más difícil de todos*

24 %

**pensó que el número “ya no debía salir”**  
*la falacia del apostador, viva*



***Ningún promedio de notas me habría dicho esto. El patrón de errores, sí.***

Concepciones documentadas en  
Kahneman y Tversky · Konold · Batanero

# ¿Cómo sé si un distractor sirve?

*Si nadie eligió una opción, no es un distractor: es relleno.*

## Tres pasos, con una hoja de cálculo

- 1 Cuente qué porcentaje eligió cada opción.
- 2 Separe al grupo con mejor y peor desempeño.
- 3 Mire si el distractor atrae más a los de menor desempeño.

### Lo que encontré

**79 % / 0 %**

El mejor distractor: lo eligió el 79 % del grupo bajo y nadie del grupo alto.

**0 %**

Dos distractores que no eligió ningún estudiante: había que reemplazarlos.

### La opción D

Inoperante en muchas de las preguntas.

*No hace falta software especializado: basta con exportar las respuestas y contar.*

# Lo que escribieron al equivocarse

Una reflexión que se queda en el dato

*“porque 22 es mayor que 20”*

Otra que coordina los dos parámetros

*“Me fijé solo en el promedio y no en la desviación de 10.  
Si la desviación es alta unas plantas crecen mucho y  
otras poco, entonces no sirve para exportar.”*

Dos fertilizantes se probaron en 50 plantas cada uno.

- A: crecimiento promedio 20 cm, desviación 2 cm.
- B: promedio 22 cm, desviación 10 cm.

Un agricultor necesita producción uniforme para exportación. ¿Qué fertilizante debería elegir y por qué?

- A. El B, porque tiene un promedio mayor que es de 22 cm.
- B. El A, porque aunque crece un poco menos, sus resultados son más consistentes y uniformes.
- C. El B, porque su desviación estándar alta indica que algunas plantas crecieron muchísimo.
- D. Cualquiera, porque 2 cm de diferencia en el promedio no es significativo.

**El mismo ítem, dos formas de razonar. Escribir el error hace visible la diferencia.**

Fragmentos anonimizados de las reflexiones escritas dentro del recurso.

# Lo que no funcionó...



## Escribir no es pensar

La reflexión era obligatoria al fallar, pero bastaba escribir cualquier cosa: aparecieron 'ns', 'no sé', letras sueltas

*Exigiré una extensión mínima en el próximo ciclo*

*Comparto esto porque una innovación que solo muestra sus aciertos no le sirve a todos.*

# Dos trampas al escribir preguntas

1

## El distractor que se delata solo

Mi opción decía que la moda era \$1.200.000 cuando era \$1.000.000: el error aritmético la volvía descartable sin pensar. Se corrigió para el segundo ciclo y algunos estudiantes la marcaron.

La tabla muestra los sueldos de 5 empleados.

Nómina Mensual - Microempresa

| Empleado             | Sueldo Mensual |
|----------------------|----------------|
| Empleado 1           | \$1.000.000    |
| Empleado 2           | \$1.000.000    |
| Empleado 3           | \$1.000.000    |
| Empleado 4           | \$1.200.000    |
| Empleado 5 (Gerente) | \$15.000.000   |

¿Cuál medida de tendencia central representa mejor lo que gana la mayoría y por qué?

- A. La media, porque incluye el sueldo del gerente y es más precisa.
- B. La moda, porque \$1.200.000 es el valor intermedio entre el mínimo y el máximo.
- ✓ C. La mediana, porque el valor central (\$1.000.000) no se ve afectado por el sueldo extremo del gerente.
- D. El rango, porque muestra la diferencia entre el que más y el que menos gana.



La tabla muestra los sueldos de 5 empleados.

Nómina Mensual - Microempresa

| Empleado             | Sueldo Mensual |
|----------------------|----------------|
| Empleado 1           | \$1.000.000    |
| Empleado 2           | \$1.100.000    |
| Empleado 3           | \$1.200.000    |
| Empleado 4           | \$1.300.000    |
| Empleado 5 (Gerente) | \$15.000.000   |

¿Cuál medida de tendencia central representa mejor lo que gana la mayoría y por qué?

- A. La media, porque utiliza los salarios de todos los empleados.
- B. La moda, porque es el sueldo que más se repite.
- ✓ C. La mediana, porque el valor central (\$1.200.000) no se ve afectado por el sueldo extremo del gerente.
- D. Cualquiera de las tres, porque todas son medidas de tendencia central válidas.



# Dos trampas al escribir preguntas

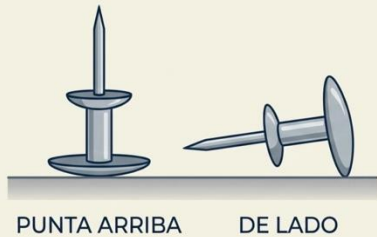
2

## El error que nadie comete

Puse un distractor con un denominador que no aparecía en el enunciado. Quizá ningún estudiante lo hubiese marcado: no era una concepción, era relleno.

Una tachuela cayó «punta arriba» 350 veces de 500 lanzamientos.

¿Cómo puede caer una tachuela?



Registro del experimento

| Resultado    | Número de veces |
|--------------|-----------------|
| Punta Arriba | 350             |
| De Lado      | 150             |
| Total        | 500             |

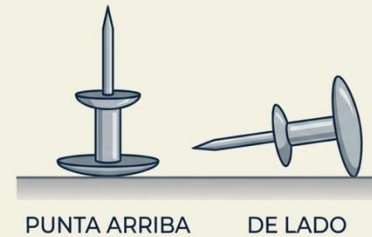
Si la lanzas una vez más, ¿cuál es la mejor estimación de la probabilidad de que caiga «punta arriba»?

- A. 50 % (0,5), porque solo hay dos opciones posibles.
- ✓ B. 70 % (0,7), porque es la frecuencia relativa observada en el experimento.
- C. 35 % (0,35), porque corresponde a 350 dividido entre 1000.
- D. 100 % (1,0), porque «punta arriba» es la moda.



Una tachuela cayó «punta arriba» 350 veces de 500 lanzamientos.

¿Cómo puede caer una tachuela?



Registro del experimento

| Resultado    | Número de veces |
|--------------|-----------------|
| Punta Arriba | 350             |
| De Lado      | 150             |
| Total        | 500             |

Si la lanzas una vez más, ¿cuál es la mejor estimación de la probabilidad de que caiga «punta arriba»?

- A. 50 % (0,5), porque solo hay dos formas posibles de caer.
- ✓ B. 70 % (0,7), porque es la frecuencia relativa observada en el experimento.
- C. 30 % (0,3), porque 150 de los 500 lanzamientos cayeron de lado.
- D. No se puede estimar, porque cada lanzamiento es independiente del anterior.



# Cuatro decisiones que se pueden replicar

1

## Escriba el error, no solo la respuesta

Antes de redactar las opciones, anote qué concepción errónea quiere capturar con cada una.

2

## Un solo intento, retroalimentación generosa

Si pueden reintentar con nota, adivinan. Separe el resultado del aprendizaje.

3

## Pida una línea escrita al fallar

“¿Qué te hizo elegir esa opción?” — en el momento, no al final de la clase.

4

## Mire el patrón, no el promedio

¿Hacia qué opción se fue el curso? Ahí está su próxima clase.

**Nada de esto exige programar ni comprar software. Exige decidir que el error merece tiempo de clase.**

# Cómo validar sus preguntas sin ser psicómetra

Pida a tres o cuatro colegas que califiquen cada pregunta, del 1 al 4, en tres criterios:

## Relevancia

¿Evalúa algo que de verdad importa del tema?

4

colegas revisaron  
mis 15 preguntas

## Claridad

¿Se entiende sin ambigüedad para el nivel del curso?

0,80

el umbral de acuerdo  
que se suele exigir

## Coherencia

¿Corresponde al estándar y el distractor captura el error previsto?

0,97

el acuerdo que  
obtuvo el instrumento

**Lo más revelador: mis colegas expertos ayudaron a afinar los distractores y los datos del segundo ciclo mostraron que la mayoría realmente funcionaron.**

# Referencias principales

## ● El error y el aprendizaje

Borasi, R. (1996). Reconceiving mathematics instruction: A focus on errors. *Ablex*.

Kapur, M. (2016). Examining productive failure... *Educational Psychologist*, 51(2), 289–299.

Sinha, T. y Kapur, M. (2021). When problem solving followed by instruction works. *Review of Educational Research*, 91(5), 761–798.

## ● Concepciones sobre el azar

Kahneman, D. y Tversky, A. (1972). Subjective probability. *Cognitive Psychology*, 3(3), 430–454.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59–98.

Batanero, C. y Serrano, L. (1999). The meaning of randomness for secondary school students. *JRME*, 30(5), 558–567.

## ● Retroalimentación y metacognición

Hattie, J. y Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.

Zimmerman, B. J. (2002). Becoming a self-regulated learner. *Theory into Practice*, 41(2), 64–70.

## ● El distractor como diagnóstico

Sadler, P. M. (1998). Psychometric models of student conceptions in science. *JRST*, 35(3), 265–296.

Treagust, D. F. (1988). Development and use of diagnostic tests... *IJSE*, 10(2), 159–169.

Shute, V. J. y Ventura, M. (2013). *Stealth assessment*. MIT Press.

# Lo que me gustaría discutir con ustedes

- ¿Cómo lograr que la reflexión escrita sea genuina y no un trámite?
- ¿Vale la pena el tiempo de diseñar distractores con intención?
- ¿Cómo llevarían esta idea a su propio tema y nivel?

---

## Gracias

Héctor Enrique Banquez Buendía · Cartagena de Indias, Colombia

bahecbu@gmail.com

<https://www.linkedin.com/in/hectorbanquez/>

Proyecto: [www.didaxislab.com](http://www.didaxislab.com)

Sitio web con simuladores interactivos de matemáticas

