

DISEÑO EXPERIMENTOS HCI

Page 1

5/20/22

Referencias:

- Research Methods in Human-Computer Interaction. Lazar, Feng and Hochheiser – capítulo 2 y 3
- Human-computer interaction: an empirical research perspective. I. Scott Mackenzie, capítulo 5.

PRÓX SEMANA

Lunes: clases

Viernes: charla invitada: Multimodal Interaction for Immersive Analytics using Augmented Reality Headsets



- Francisco Ortega, Colorado State University
- Director Nuilab <https://nuilab.org/>

DESCRIPTIVE INVESTIGATIONS (CLASE ANTERIOR...)

Describir

- Observations
- Surveys
- Focus Groups

“X está pasando”

[Lazar, Feng, Hochheiser 2010]

BEHAVIORAL INVESTIGATIONS

Relaciones entre varios factores

- Observations
- Field studies
- Surveys

“X está relacionado a Y”

[Lazar, Feng, Hochheiser 2010]

EXPERIMENTAL INVESTIGATIONS

Para determinar efecto causal entre dos factores

- Experimentos controlados

"X es la causa de Y"

[Lazar, Feng, Hochheiser 2010]

EXPERIMENTOS

Tratamientos (treatments)

- Técnicas, dispositivos, procedimientos que queremos comparar

Unidades (units)

- Objeto a quien aplicamos los tratamientos
- E.g. personas de cierto género, edad, habilidades digitales

Método de asignación (assignment method)

- Cómo asignamos tratamientos a unidades

[Lazar, Feng, Hochheiser 2010]

EXPERIMENTOS

Ventaja: internal validity

Desventaja: riesgo de low external validity

[Gergle & Tan 2014]

| | | | | | | | | | | | | | |
|----------------|------------|--------|--------|---------|--------|--------|--------|--------|--------|------------|--------|------------|----------------|
| ~ ` | ! 1 | @ 2 | # 3 | \$ 4 | % 5 | ^ 6 | & 7 | * 8 | (9 |) 0 | - _ | + = | ← Backspace |
| Tab ⇐ ⇒ | Q | W | E | R | T | Y | U | I | O | P | { [| }] | \ |
| Caps Lock ⬆ | A | S | D | F | G | H | J | K | L | : | " " | ↵ Enter | |
| Shift ⬆ | Z | X | C | V | B | N | M | < , | > . | ? / | | Shift ⬆ | |
| Ctrl | Win Key | Alt | | | | | | | Alt | Win Key | Menu | Ctrl | |

| | | | | | | | | | | | | | |
|----------------|------------|--------|--------|---------|--------|--------|--------|--------|--------|------------|--------|------------|----------------|
| ~ ` | ! 1 | @ 2 | # 3 | \$ 4 | % 5 | ^ 6 | & 7 | * 8 | (9 |) 0 | { [| }] | ← Backspace |
| Tab ⇐ ⇒ | " " | < , | > . | P | Y | F | G | C | R | L | ? / | + = | \ |
| Caps Lock ⬆ | A | O | E | U | I | D | H | T | N | S | - _ | ↵ Enter | |
| Shift ⬆ | : | Q | J | K | X | B | M | W | V | Z | | Shift ⬆ | |
| Ctrl | Win Key | Alt | | | | | | | Alt Gr | Win Key | Menu | Ctrl | |

[Lazar, Feng, Hochheiser 2010]

SIGNIFICANCIA ESTADÍSTICA

“Mike mide 1,8metros. Mary mide 1,67metros. Por lo tanto, Mike es más alto que Mary”

“La altura promedio de tres hombres (Mike, John y Ted) es 1,68metros. La altura promedio de tres mujeres (Mary, Rose y Jessica) es 1,78metros. Por lo tanto, las mujeres son más altas que los hombres”

[Lazar, Feng, Hochheiser 2010]

SIGNIFICANCIA ESTADÍSTICA

H0: Defendant is innocent

| | | Jury decision | |
|---------|------------|---------------|--------------|
| | | Not guilty | Guilty |
| Reality | Not guilty | ✓ | Type I error |
| | Guilty | Type II error | ✓ |

Table 2.3 Type I and Type II errors in the judicial case.

[Lazar, Feng, Hochheiser 2010]

SIGNIFICANCIA ESTADÍSTICA

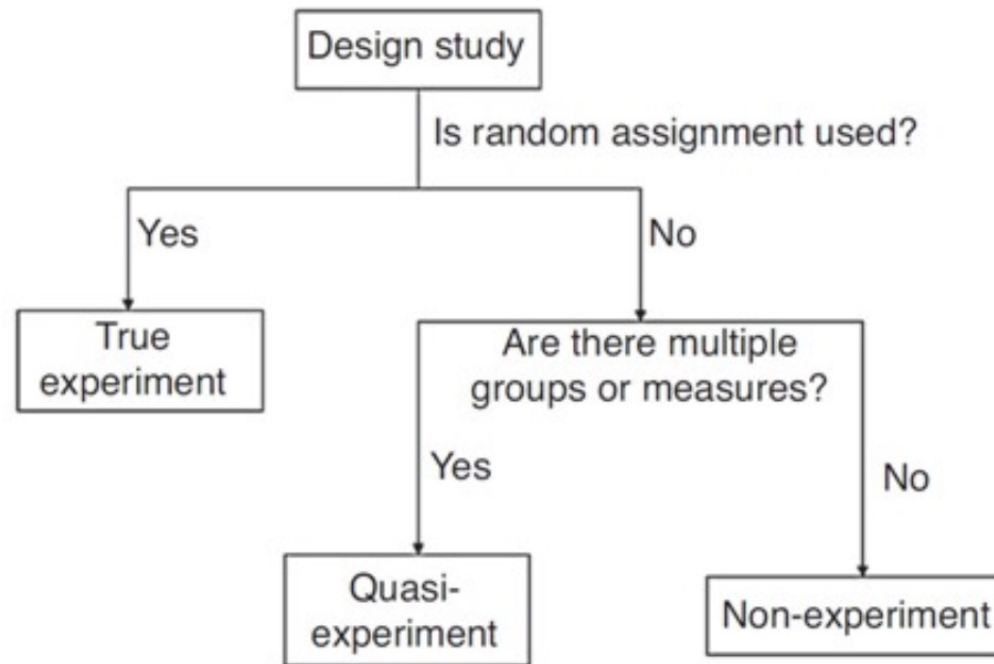
H0: There is no difference between ease of use of ATMs with touchscreen or buttons.

| | | Study conclusion | |
|---------|----------------------------------|------------------|----------------------------------|
| | | No difference | Touchscreen ATM is easier to use |
| Reality | No difference | ✓ | Type I error |
| | Touchscreen ATM is easier to use | Type II error | ✓ |

Table 2.4 Type I and Type II errors in a hypothetical HCI experiment.

[Lazar, Feng, Hochheiser 2010]

EXPERIMENTOS



[Lazar, Feng, Hochheiser 2010]

EXPERIMENTAL RESEARCH IN HCI

Mostrar cómo la manipulación de una variable tiene una influencia causal en otra variable de interés

Ventajas: internal validity; posibilidad de replicación; datos cuantitativos que se analizan con estadística

Limitaciones: requiere hipótesis bien definida y demostrable con conjunto de variables; difícil si hay muchos factores que influyen; low external validity; settings de laboratorio muy artificiales

[Gergle & Tan 2014]

PERO: USABILITY EVALUATION CONSIDERED HARMFUL?

Nos gustan los experimentos controlados

Esto, tiende a eliminar ideas demasiado temprano y empujarnos a ir hacia lo que podemos medir

La visión es “critical”; puede permitir lo que somos capaces de hacer; no necesariamente es un producto rápidamente; es difícil entender cómo se adoptarán; para innovaciones la evaluación puede ser “meaningless”

Podemos hacer (en casos justificados) “a design rationale, a vision of what could be, expected scenarios of use, reflections, case studies, participatory critique, and so on”

Elegir un método que efectivamente nos informe, para lo que estamos estudiando

[Greenberg & Buxton 2008]

VARIABLE EXPERIMENTAL



VARIABLE

Variable independiente

- Variable manipulada por investigadores
- Debe haber variación bien controlada

Variable dependiente

- Medida de outcome cuyo valor se predice que variaría basado en los niveles de la variable independiente
 - Ejemplo: número de clics, tiempo de respuesta, movimiento de ojos

VARIABLE INDEPENDIENTE O FACTOR

Típicamente una propiedad de una interfaz

- Dispositivo (Mouse, stylus, ...)
 - Modalidad de feedback (audio, visual, táctil)
 - Tamaño de display
 - Visualización de árbol (tradicional, lista, multi columna)
 - etc etc
-
- Sistemáticamente dar una interfaz y luego otra!

[Scott Mackenzie 2013]

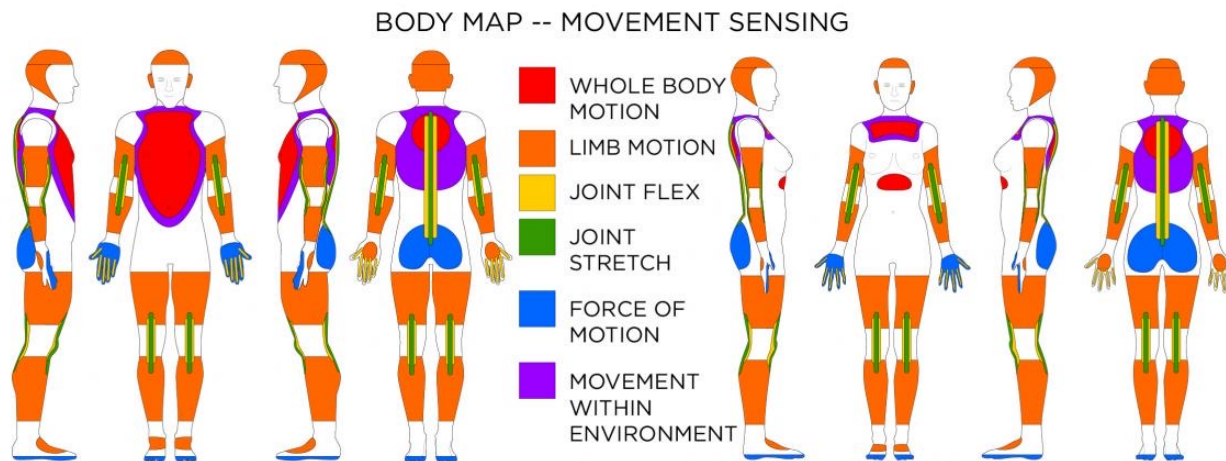
<https://www.youtube.com/watch?v=-q702OTBpT8> . Leer: <https://hal.inria.fr/hal-01144163/document>

VARIABLE INDEPENDIENTE O FACTOR

También puede ser otras cosas:

- Edad de usuarios
- Género
- Mano (zurdos, diestros)
- Nivel de competencias digitales
- Ruido ambiente, luz, ...
- Religión, idioma, altura, color de pelo, etc etc etc!

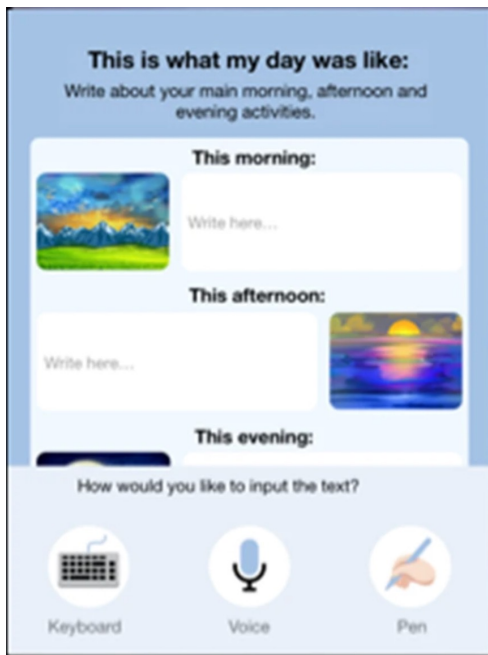
[Scott Mackenzie 2013]



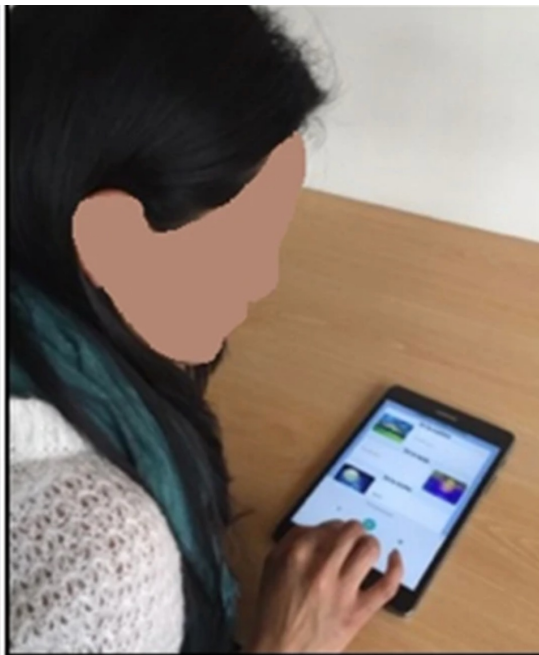
VARIABLE INDEPENDIENTE O FACTOR

¿Cómo especificar?

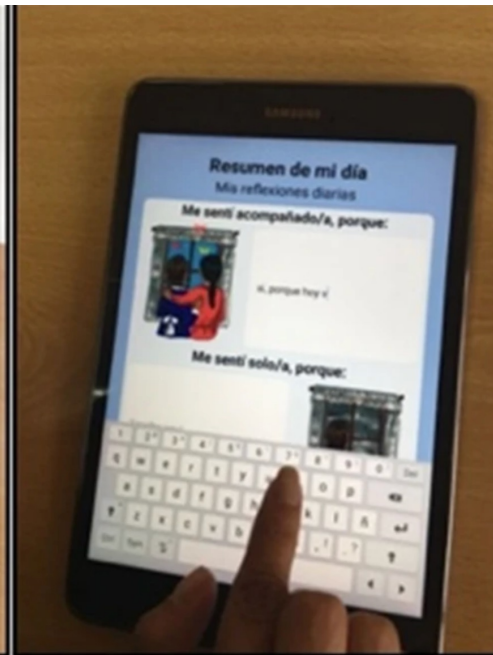
- Característica + nivel elegido para la prueba



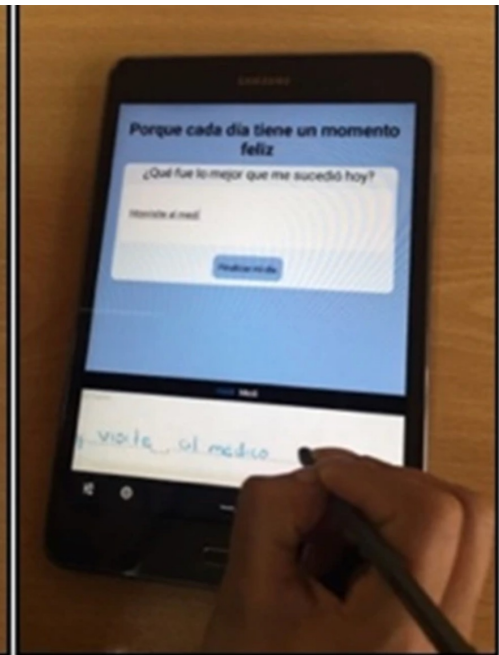
(a)



(b)



(c)



(d)

VARIABLE INDEPENDIENTE O FACTOR

Puede haber más de una por estudio.

Pero....

VARIABLE INDEPENDIENTE O FACTOR

Puede haber más de una por estudio.

Pero....

“To take a toy example, suppose you wanted to establish a link between chocolate and baldness. You could then get a group of 10,000 men (a pretty big sample size by all accounts) to report on their consumption of M&Ms, Twix and Mars Bars over a period of time. In addition, you record the rate of going bald in the group over time.

Once you have your chocolate and baldness data, you run tests on everything you can think of. Do men who eat only M&Ms go bald younger? Do young men who eat both Mars and M&Ms but not Twix go bald on top more often than the front? Do older unmarried men who don't exercise and eat none have a lower incidence of baldness?”

<https://scienceinthenewsroom.org/resources/statistical-p-hacking-explained/>

DEPENDENT VARIABLE

Típicamente:

- Velocidad
- Precisión

Otras?

- Tiempo de preparación, tiempo de acción, cambios en vista, transiciones entre mouse y teclado, etc etc (cualquier aspecto de comportamiento humano que sea **medible!**)

MÁS VARIABLES

Variables de control

Variables random

Variables de confusión (confounding)

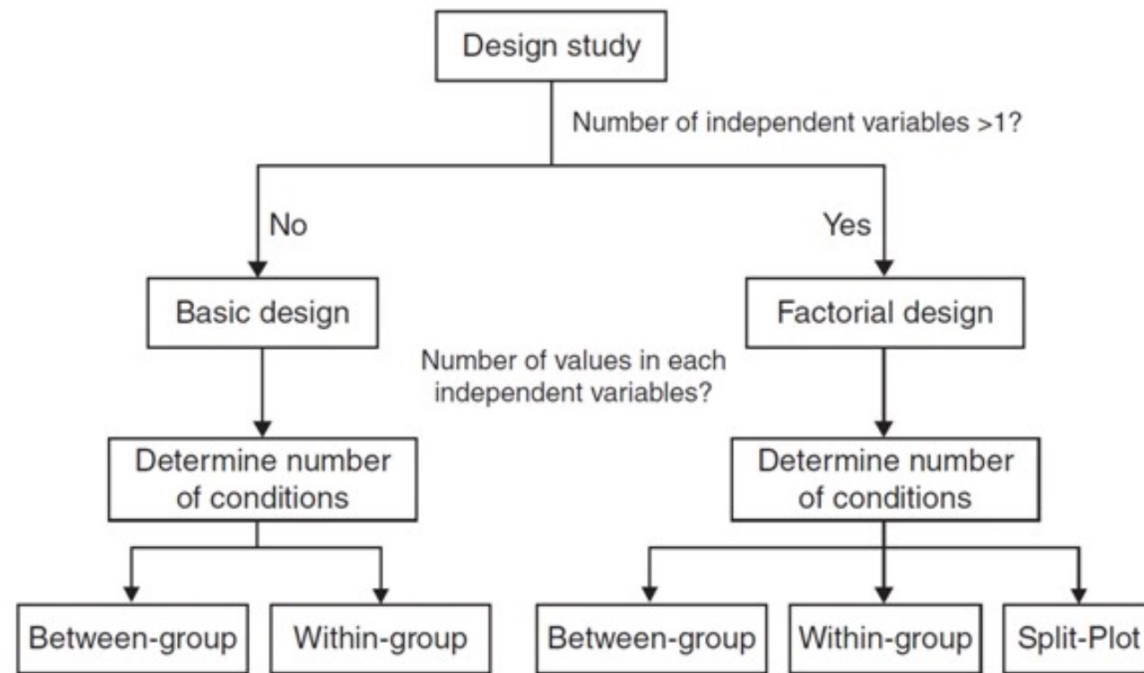
¿CUÁNTOS PARTICIPANTES?

Ideal: similar a estudios previos

Demasiados => significancia estadística pero puede ser con efecto pequeño

Pocos => difícil llegar a significancia estadística (aunque pueda existir!)

DISEÑO DE ESTUDIOS



[Lazar, Feng, Hochheiser 2010]

Within group = within subjects, o repeated measures

WITHIN 0 BETWEEN?

Ventajas y desventajas?

| Participante | Test Condition | | |
|--------------|----------------|---|---|
| 1 | A | B | C |
| 2 | A | B | C |

| Participante | Test Condition |
|--------------|----------------|
| 1 | A |
| 2 | A |
| 3 | B |
| 4 | B |
| 5 | C |
| 6 | C |

[Scott Mackenzie 2013]

WITHIN 0 BETWEEN?



INTERFERENCIA EN WITHIN-SUBJECTS

Es posible ir aprendiendo mientras vamos pasando de una condición a otra (*practice/learning effect*)

También es posible que los participantes se cansen (*fatigue effect*).

¿Qué hacer?

INTERFERENCIA EN WITHIN-SUBJECTS

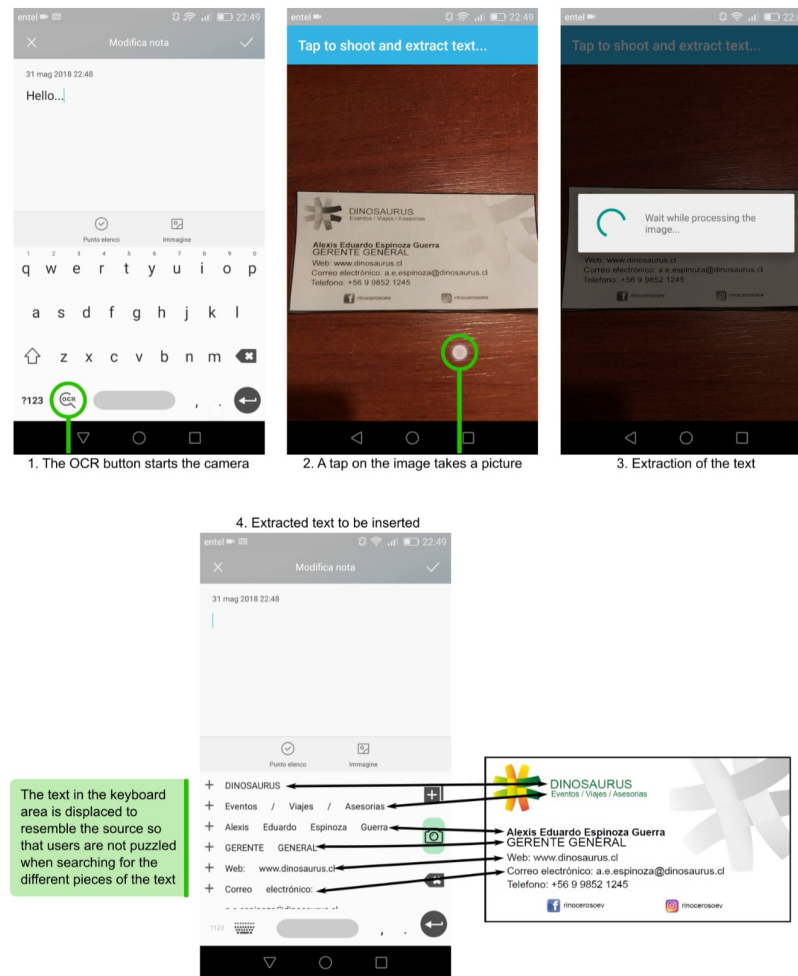


FIGURE 1. CameraKeyboard use process. At the beginning, users can write through the QWERTY on-screen keyboard as usual. When users need to start the camera, they tap on the OCR button (1), then pointing the camera to the text of interest, user can take a picture (2). The image is sent to the Google Cloud Vision API and the text is extracted (3). Once the text is returned to the smartphone, it is displaced similarly to the source and user can insert the entire text, a single line, or a single word.

INTERFERENCIA EN WITHIN-SUBJECTS



FIGURE 3. The samples used for the user studies: business cards (top-left), document IDs (top-center), lottery cards (top-right), papers (bottom). Although texts are different among comparable samples, they have the same number of character and special character for any of the data to be transcribed.

INTERFERENCIA EN WITHIN-SUBJECTS

Es posible ir aprendiendo mientras vamos pasando de una condición a otra (*practice/learning effect*)

También es posible que los participantes se cansen (*fatigue effect*).

¿Qué hacer?

* Counterbalancing => Latin Squares

| | |
|---|---|
| A | B |
| B | A |

| | | |
|---|---|---|
| A | B | C |
| B | C | A |
| C | A | B |

| | | | |
|---|---|---|---|
| A | B | C | D |
| B | C | D | A |
| C | D | A | B |
| D | A | B | C |

[Scott Mackenzie 2013]

INTERFERENCIA EN WITHIN-SUBJECTS

Es posible ir aprendiendo mientras vamos pasando de una condición a otra (*practice/learning effect*)

También es posible que los participantes se cansen (*fatigue effect*).

¿Qué hacer?

* Counterbalancing => Latin Squares => Balanced Latin Squares

| | |
|---|---|
| A | B |
| B | A |

| | | |
|---|---|---|
| A | B | C |
| B | C | A |
| C | A | B |

| | | | |
|---|---|---|---|
| A | B | C | D |
| B | C | D | A |
| C | D | A | B |
| D | A | B | C |

| | | | |
|---|---|---|---|
| A | B | D | C |
| B | C | A | D |
| C | D | B | A |
| D | A | C | B |

* O: randomizing

[Scott Mackenzie 2013]

ESTUDIOS LONGITUDINALES

Cuando la investigación quiere enfocarse en aprendizaje o adquisición de habilidad

Se testea por periodos prolongados de tiempo y se mide mejoras en performance

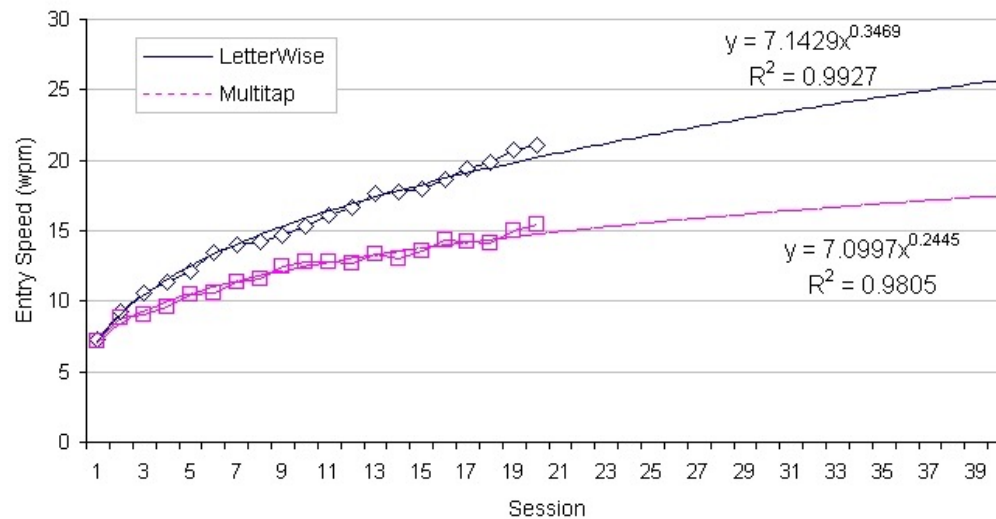
En vez de eliminar aprendizaje, queremos medirlo!

ESTUDIOS LONGITUDINALES



¿Cómo escribo HOLA?

¿Cómo escribo ON?





PREGUNTAS?