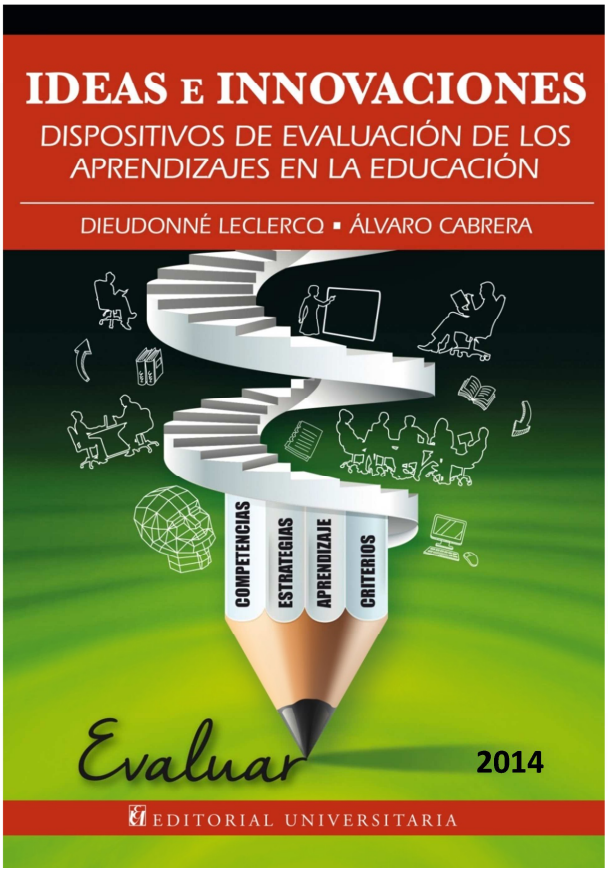

Dieudo LECLERCQ

Álvaro CABRERA MARAY

Universidad de Chile

www.universitaria.cl

Directores de la publicación:

Dieudonné Leclercq
Universidad de Liège (ULg)

Álvaro Cabrera Maray
Universidad de Chile (UCH)

IDEAS e INNOVACIONES
Innovaciones en Dispositivos de Evaluación de los Aprendizajes en la enseñanza Superior
2014

Se pueden bajar gratuitamente
desde http://orbi.uliege.be, después Leclercq D., o
desde www.evaluaraprendizajes.cl
-Los **resúmenes** de los 23 capítulos
del libro IDEAS http://hdl.handle.net/2268/173543
-El **índex** de este libro para buscar entre
entradas de 1500 conceptos y
400 de autores http://hdl.handle.net/2268/180060

Dieudonné Leclercq

Dr. en Educación (1975) en « La Metacognición vía la autoevaluación con grados de certeza » y con postdoctorales en las universidades de Pittsburgh y UCLA. Fue profesor en las Universidades de Namur (1975-1980) y de Liège (1980-2010). Es emérito desde 2010. Ensena como invitado en las Ues. de Liège y Paris 13. Recibió el título de *Honorary Member of the World Cultural Council* (México). Ha colaborado, en Chile, con la U de Chile (UCH -Santiago), la UMCE, la UCT (Temuco), la UC del Maule, la UNAB y la UCSC (Concepción). En Perú con la PUCP y el SINEACE (Lima), la UNSAAC (Cusco) y la UNTRM (Chachapoyas). En México con la U A Chapingo. En España con la U de Sevilla y la U de Deusto (Bilbao). d.leclercq@uliege.be

Álvaro Cabrera Maray

Licenciado en Artes mención Teoría de la Música, y Master en Pedagogía en Educación Superior de la U. de Liège (Bélgica). Ha sido profesor en la Facultad de Artes y en Cursos de formación General, trabajando en el Depto. Estudios de Pregrado de la U. de Chile a cargo del Área de Formación. Integró la Red nacional de Centros de Enseñanza-Aprendizaje y la de expertos SCT-Chile sobre sistema de créditos transferibles. Trabajaba en el Ministerio de Educación de Chile, coordinando los programas de la reforma educacional en Educación Superior. alvarocabreramaray@gmail.com

Contenidos del libro IDEAS:

ES: Calificación ; Evaluación ; Productos ; Meta-cognición ; Resolución de problemas ; Proyectos ; Trabajo de grupo ; Portafolio ; Vigilancia cognitiva ; Pruebas de Progreso ; Taxonomía de Bloom ; Auto-evaluación ; Grados de certeza ; Test de Concordancia de Script ; Retroinformación ; calidades ; validez

EN : Assessment ; Evaluation ; Outcomes ; OSCE ; MCQ ; PARMs ; Metacognition ; Problem solving ; Projects ; Group produced work ; Portfolio ; Cognitive vigilance ; Progress Tests ; Bloom's Taxonomy ; Self-assessment ; Confidence Degrees ; Concordance Script Test ; Feedbacks ; Edumetrics ; Metacognitive Spectral Test ; ETIC PRAD ; quality ; validity

FR : Notation ; Evaluation ; Résultats ; ECOS ; QCM ; PARMs ; Métacognition ; Résolution de problèmes ; Projets ; Travail de groupe ; Portfolio ; Vigilance cognitive ; Tests de progression ; Taxonomie de Bloom ; Auto-évaluation ; Degrés de certitude ; Test de Concordance de Script ; Rétro-information ; Edumétrie ; Test Spectral Métacognitif ; qualités d'une évaluation ; validité d'une mesure

IDEAS = **I**nnovaciones en **D**ispositivos de **E**valuación de los **A**prendizajes en la educación **S**uperior

La lista de los capítulos y el resumen de caca uno

aparece a continuación después de este capítulo.

Capítulo XXIII

Índices Cuantitativos en Docimología

Dieudonné Leclercq, Álvaro Cabrera y Rodrigo Roco

Parte 1. Índices cuantitativos: conceptos y debates acerca de su uso

A. ¿Qué son y para qué sirven los índices cuantitativos?

Los índices cuantitativos son síntesis de observaciones expresadas en forma numérica y que sirven para ayudar al profesor y al estudiante a *reflexionar* y *decidir.*

Tabla 1: Tipos de índices cuantitativos en docimología[206]

Cuando los índices	...ver sección
...resumen una gran cantidad de datos y permiten mostrar uno o más patrones o tendencias (centrales, de dispersión, de distancia, etc.).	E, F, G
...permiten comparar resultados obtenidos con resultados anteriores (por ejemplo, en una situación de pre-test / post-test para los mismos estudiantes con las mismas preguntas).	K, L, M
...permiten situar una clase, un estudiante, etc., en comparación con las referencias establecidas de antemano (por ejemplo, los resultados de otras clases o de otros países).	H, I
...permiten verificar si se cumplen ciertos criterios (por ejemplo, umbrales de logro para un estudiante, calidad de una pregunta, fiabilidad de una prueba o test), o se refuerza o refuta una hipótesis.	J, N, O, P

Estos índices resumen o combinan varios *datos de base,* varias evaluaciones (puntajes, comparaciones) de:

- un estudiante → para tomar decisiones con propósitos formativos o certificativos, conforme a reglas del docente o del tribunal escolar. Los datos pueden provenir de un mismo profesor o de varios profesores.
- un grupo de estudiantes → para tomar decisiones sobre un curso, o sobre una prueba y sus ítems (preguntas, criterios).

[206] Entendiendo la docimología como la ciencia de los exámenes y de la atribución de puntajes a los estudiantes.

Tabla 2: Tipos de índices cuantitativos

Tipos de índice	Lista de índices o instrumentos (y secciones del capítulo)
De tendencia central	La Moda (E1), la Mediana (F2), la Media (G3), y, si se usan grados de certeza[205], la Confianza Media y la Imprudencia Media.
De dispersión / diversidad vs. monopolización / concentración	La Variedad (E2), la Entropía y la Neguentropía (E3), el Rango Intercuartiles[206] (F3), la Desviación Estándar (G4)
Gráficos	De torta (E), de barras (E), histogramas (G2), polígonos de frecuencia (G2), hemispectros dobles (G7), dispersión o nube de puntos (Q y R), relación de tendencia lineal o curvilineal (P)
De forma de la distribución o curva	En i, en J, en U, en Gauss (G), asimetría o SKEWNESS (G5)
De progresión (de evolución)	La Ganancia (o progreso), la Amplitud de Efecto (M), la Ganancia Relativa (L)
De posición del resultado individual en el grupo	El orden (H), la nota o puntaje z (I1), el decil (I2), el centil o percentil (I2)
Límites de confianza	Determinación de los Límites de Confianza (O)
Sobre preguntas o ítems	Facilidad / dificultad (P), discriminación por r_{pbis} (P3)

Estos índices cuantitativos también tienen *restricciones*. Algunas de ellas son:

- Estos índices no caracterizan al estudiante, sino que a sus resultados en un momento particular.
- Estos índices no se pueden utilizar con todos los tipos de datos (ver secciones E, F y G).
- Estos índices pueden (y en algunos casos deben) ser expresados con su "error de medición".
- Algunos índices de posición (por ejemplo, el puntaje *z*) necesitan conocer previamente otros índices de la prueba (como la Media y la desviación estándar) para ser calculados.

B. Tipos de datos: tres categorías principales

B.1. Categorías Nominales

La *evaluación*, que consiste en *medir* + *juzgar*, necesita estar referida a criterios y umbrales de logro. Existen categorías *nominales* que no se pueden ordenar y que, a menudo, *no sirven para emitir juicios*. Por ejemplo, en el *sistema educativo* no pueden ser criterios de éxito, fracaso o restricción:

[207] Ver Capítulo 17.

[208] Se habla de cuartiles cuando una distribución de notas es dividida en 4 grupos: el 25% con mejores notas, el 25% siguiente, y así hasta el 25% con peores resultados.

- *la nacionalidad*, aunque sí lo puede ser en otro ámbito, como por ejemplo en el sistema electoral: el Presidente de la República debe poseer la nacionalidad del país que representa.
- *el género*, aunque en los deportes muchas veces las mujeres participan en competencias diferentes a las de los hombres.
- *el estilo de pensamiento* predominante (convergente o divergente), aunque lo puede ser en otros ámbitos, como en algunas empresas que prefieren personas divergentes (de publicidad, por ejemplo) mientras que en otras prefieren personas convergentes (bancos, por ejemplo).
- *la edad*, aunque sí lo puede ser, por ejemplo, en el acceso a juegos peligrosos o espectáculos agresivos o considerados inmorales.
- *el ciclo diario* (como estudiar durante la noche o en la madrugada), aunque sí puede ser restrictivo en otro ámbito, como por ejemplo para trabajar en una empresa como vigilante nocturno.

Existen otras caracterizaciones nominales que sí pueden servir para juzgar. En el sistema educacional, por ejemplo: el plagio, la peligrosidad para los otros, la trampa (en exámenes), la toxicidad.

Por otro lado, los estudiantes son en sí mismos categorías nominales. No hay un orden natural (*a priori*) entre ellos. Es por eso que cuando los representamos en un gráfico no dibujamos una línea quebrada o un polígono de frecuencias (ver secciones F y G), puesto que ello daría la impresión de una evolución o de fenómenos de ascenso y descenso, que se deberían únicamente a la posición de los estudiantes en el gráfico. Sin embargo, dicha posición es aleatoria (por ejemplo, el orden alfabético de sus nombres, o el código que le asignamos a cada uno de ellos: e1, e2, e3, etc.).

B.2. Categorías Ordinales

Consideremos la siguiente rúbrica y el número de jueces (9) que han atribuido un determinado nivel al desempeño (performance) de un estudiante ficticio:

Tabla 3: Distribución de las notas de 9 jueces para un mismo estudiante

1. Insuficiente	2. Débil	3. Aceptable	4. Satisfactorio	5. Bien	6. Excelente
0	2	2	1	3	1

Aunque están codificadas con números (del 1 al 6) estas notas son categorías ordinales (puesto que el orden no puede ser cambiado) pero NO son métricas. En consecuencia, los datos que se muestran en la Tabla 3 permiten calcular los siguientes índices edumétricos[209]:

[209] De acuerdo con la definición de Carver (1974) y sus dos dimensiones de los tests: Psicométrica y Edumétrica.

- la *Moda*[210]: que aquí es "5. Bien" (el nivel de la rúbrica más frecuentemente asignado por los nueve jueces).
- la *Mediana*[211]: que aquí es "4. Satisfactorio" (la elección del juez que está justo en el valor que separa las notas en dos mitades iguales y que en este ejemplo muestra 4 notas por debajo y 4 notas por arriba).

Sin embargo, en principio, no se puede calcular la *Media*[212] porque no tenemos garantía de que las distancias entre los intervalos son iguales. Algunos autores, sin embargo, calculan de todas maneras la Media (y la Desviación Estándar) en escalas ordinales como la presentada aquí. En la actualidad, existe un debate sobre este punto.

B.3. Categorías Métricas de intervalos y puntajes Z

A) CATEGORÍAS ENGAÑOSAMENTE MÉTRICAS

Consideremos la siguiente rúbrica (que describe la cantidad de errores en 5 pruebas), y la cantidad de veces que cada situación propuesta se presentó para un mismo estudiante:

Tabla 4: Distribución de errores de un estudiante en cinco pruebas

1. Menos de 3 errores	2. Entre 3 y 7 errores	3. Entre 7 y 12 errores	4. Más de 12 errores	TOTAL (pruebas)
1	2	2	0	5

Lo primero que podemos observar es que los intervalos entre las cuatro categorías no son iguales. Por esa razón no se puede calcular la Media sobre la base de los códigos asignados (1, 2, 3 y 4). Una variable de tipo continuo (el número de errores) ha sido policotomizada (agrupada en varias categorías). Entonces, si realmente queremos calcular la Media se debería indicar cada vez el número exacto de errores, tal como se muestra en la Tabla 5:

Tabla 5: Número de errores de un estudiante en cinco pruebas

Prueba	A	B	C	D	E	Total
Número de errores	8	4	11	1	5	29

[210] La *Moda* es aquel valor de una distribución que se repite o aparece la mayor cantidad de veces.

[211] La *Mediana* corresponde a aquel valor de una distribución (o conjunto de valores) que la divide en dos partes iguales. La Mediana se ubica en el centro del conjunto cuando los números se ordenan de menor a mayor.

[212] La *Media* (también conocida como media aritmética, promedio, esperanza o valor esperado) se define como el valor de una distribución que se obtiene al sumar todos los valores que la componen y dividir dicha suma por el número de valores de esa distribución.

El número de errores por prueba sí es una variable métrica, mientras que el número de veces que un estudiante puede clasificarse en una determinada categoría de cantidad de errores ("menos de 3", "entre 3 y 7", etc.) no lo es.

B) CATEGORÍAS MÉTRICAS DE INTERVALOS IGUALES SIN "CERO ABSOLUTO" Y PUNTAJES Z (VER SECCIÓN I.1)

Consideremos una prueba con Preguntas de Selección Múltiple (PSM), donde el estudiante recibe 1 punto por cada respuesta correcta y –0,5 puntos por cada error. En este caso, existe la posibilidad de que el total de un estudiante sea negativo. Si la prueba tiene 20 preguntas, el verdadero puntaje mínimo posible no es 0 puntos sino –10 puntos (20 * –0,5 para 20 errores). Por esa razón, se puede calcular la Media, pero no se puede afirmar que un puntaje total de 10 puntos es dos veces mejor que un puntaje de 5 puntos, porque ¿qué se puede decir de la relación entre –5 y 5?

Las categorías métricas permiten situar o ubicar la nota de un estudiante de manera relativa a la Media de la distribución de las notas del grupo de estudiantes al cual este pertenece. Para ello se utiliza la fórmula de la nota o puntaje $Z = (X - M) / DE$, donde X es la nota del estudiante, M la media de los resultados del grupo y DE la desviación estándar de estos resultados. Una nota Z negativa significa "por debajo de la media" y una positiva quiere decir "superior a la media". Las notas o puntajes Z varían generalmente de –3 hasta +3.

B.4. Categorías Métricas de proporciones (con un 0 absoluto)

En este tipo de categoría el 0 (cero) es absoluto. Es lo que pasa, por ejemplo, cada vez que no es posible obtener un número negativo: número de ideas, de buenas respuestas, de intervenciones en un foro electrónico, de conexión en un sitio de discusión, etc. En estos casos, 6 es siempre dos veces más que 3. Por lo tanto, pueden calcularse proporciones.

B.5. Categorías Métricas de proporciones (%) con un máximo

Con este tipo de categoría sí se puede calcular la distancia entre la performance, desempeño o nivel de logro, con respecto al nivel máximo. La distancia que exista es lo que se puede mejorar y que se conoce como "Ganancia posible". Este tipo de razonamiento no es permisible cuando la máxima es desconocida. Por ejemplo: el número de intervenciones en un foro de discusión o el número de ideas no poseen habitualmente un número máximo conocido.

En muchas otras ocasiones la máxima sí es conocida. Por ejemplo, en una prueba de N preguntas, la máxima corresponde a un éxito en cada una de las N preguntas, de modo que se puede calcular un porcentaje o proporción. Por ejemplo, un estudiante que ha tenido éxito en 7 de las 14 preguntas de una prueba ha logrado 50% de éxito.

B.6. Tipos de datos e índices permitidos

Según el tipo de medición (prueba, cuestionario con escala de Likert, etc.) es posible considerar cuatro tipos de datos, que a su vez autorizan (o prohíben) ciertas operaciones matemáticas con cada uno de ellos. La Tabla 6 presenta una vista general de conceptos que definiremos e ilustraremos más adelante.

Tabla 6: Tipos de datos e índices permitidos

CON DATOS DE LAS CATEGORÍAS...	...nominales	..ordinales	...métricas de intervalos iguales	...métricas de proporciones	...métricas de proporciones con máxima
=, ≠	sí	sí	sí	sí	sí
>, <		sí	sí	sí	sí
+, −			sí	sí	sí
*, /				sí	sí
%					sí
Error Estándar de %	sí	sí	sí	sí	sí
Moda	sí	sí	sí	sí	sí
Mediana y Rango Semi-Intercuartiles		sí	sí	sí	sí
Media, Desviación estándar, Amplitud de Efecto			sí	sí	sí
Ganancia Relativa					sí

C. ¿Pueden los docentes y profesores hacer investigación?

C.1. ¿Qué diferencias existen entre la investigación en el aula y la investigación de laboratorio?

Esta pregunta intentará ser respondida a lo largo de todo el capítulo, a través de casos y experiencias de investigación en docencia. Conscientemente hemos escrito "de laboratorio" y no "en laboratorio", porque "el laboratorio puede ser tan grande como el mundo" como señala Benjamín Bloom[213].

[213] Bloom fundó –junto a G. de Landsheere, T. Husen, N. Postlethwaite y otros– la IEA (International Association for the Evaluation of Educational Achievement), siendo precursor del actual programa PISA de la OCDE (Program International for Student's Assesment). Bloom consideraba la diversidad de sistemas educacionales en el mundo como una vasta experimentación no voluntaria y cuyos datos pueden ser tratados vía análisis multivariados.

C.2. La investigación a partir de muestras representativas

Caso 1: La encuesta PISA en matemáticas. Este tipo de encuestas tiene por objetivo estimar (con un margen de error mínimo) los resultados de una población entera (por ejemplo, todos los jóvenes de 15 años de Chile), a partir de los resultados de un subconjunto restringido (una muestra) de estudiantes representativos de esa población

Para ser representativa, una muestra debe haber sido establecida de acuerdo con ciertas reglas. Siguiendo el ejemplo, estas reglas son:

La primera: las escuelas donde se aplicará la prueba deben ser elegidas de acuerdo con un muestreo estratificado que garantice que los diferentes tipos de escuela existentes en el país estarán representados en número suficiente a la hora de seleccionarlas al azar. Dentro de las escuelas se elegirán las clases al azar y, dentro de las clases se elegirá de manera aleatoria a los estudiantes a ser evaluados[214].

La segunda: el número de alumnos evaluados debe ser suficientemente grande (al menos 2.000). ¿Por qué? Porque para estar seguros de que el verdadero valor de un porcentaje (por ejemplo, el nivel de logro) de la población (y no solo de la muestra) está dentro de un rango de 99% de credibilidad, debemos tener en cuenta el error de medición (también expresado en %). Este índice, a diferencia de los otros tratados en este capítulo, está orientado hacia la generalización de los resultados, es decir, hacia la inferencia. En tal caso, hablamos de "estadística inferencial" que busca poder extrapolar, hacia toda la población, lo que se observa en la muestra.

En la Tabla 7 se presentan los valores que N debe alcanzar para reducir el error de medición y, por ende, los límites de confianza de un porcentaje observado en una muestra representativa cuando se quiere deducir o inferir el porcentaje real para toda la población.

No obstante, es necesario tener en cuenta dos principios:

- Mezclar los resultados de muchas clases de estudiantes permite aumentar el número; sin embargo nos hace perder en contextualización, dado que tendemos entonces a "borrar" las diferencias entre clases, las cuales pueden derivarse, por ejemplo, del origen social de los estudiantes, de los requisitos previos para estar en esas clases (selección), de los métodos del profesor de la clase, etc.
- La dispersión de los resultados a veces es más importante que la media. Si un país, por ejemplo, tiene un promedio de 73%, puede ser mucho más importante saber cuál es la proporción de estudiantes por debajo de la puntuación de 50% (es decir, muy por debajo de la media nacional), pues esa información es más relevante para la acción.

[214] Incluso si la prueba es aplicada a todos los estudiantes, solo serán analizados en las estadísticas aquellos estudiantes que han sido previamente elegidos al azar para integrar la muestra.

C.3. El Error Estándar para la Medición de un porcentaje

A) ¿CÓMO SE CALCULA EL ERROR ESTÁNDAR DE MEDICIÓN DE UN PORCENTAJE (EEM%)?

La fórmula es: $EEM\% = \sqrt{p*q/N}$

Donde:

- p es la proporción (o tasa) de éxito (por ejemplo 0,6 o 60%),
- q es la proporción complementaria, es decir: $1 - p = q$ (aquí 0,4 o 40%)
- N es el número de observaciones sobre las cuales se calculan p y q.

B) ¿PARA QUÉ SIRVE EL EEM%?

Para establecer los límites de confianza, el EEM% debe indicarse casi dos veces menos (–1,96 * EEM%) y casi dos veces más (+1,96 * EEM%)[215] que el valor del porcentaje observado en la población (en nuestro ejemplo, 60%), estableciendo así los límites de confianza, también llamados intervalo de confianza, en el cual se encontraría el valor o porcentaje real. Por lo tanto, si el porcentaje de logro observado en la muestra es del 60% y el Error de medición es del 5%, podemos decir, con un 95% de probabilidades de no engañarnos –o con un 5% de posibilidades de error– que el porcentaje de logro en la población total oscila, aproximadamente, entre 50% y 70% (60% ± (1,96 *5%)).

La Tabla 7 describe los EEM% y los límites de confianza para diferentes valores de N, considerando que $p = 0,6$ (60%) y que, por lo tanto, q (o $1 - p$) = 0,4 (40%).

La Tabla 7 muestra que:

- El aumento del tamaño de la muestra debe ser muy grande para pasar de un intervalo de incertidumbre (o de confianza) del 3% a uno del 2%.
- Nunca consideramos el tamaño de la población a la que se quiere extrapolar las inferencias que autoriza la muestra. Los valores de la Tabla 7 serán válidos para una población de 10.000 personas o una de 100 millones de personas.

C.4. Investigación en la clase: dos malas noticias

> Caso 2: Una clase de 300 estudiantes de 1er año de Universidad.
>
> En esta clase no se cumplen las condiciones de muestreo al azar: los 300 estudiantes no son representativos, porque no fueron elegidos al azar de entre la población general de su grupo de edad en su país o zona geográfica. Por lo mismo, los estudiantes de esa clase no pueden servir para una estimación estadística para toda la población estudiantil del país.

Mala noticia 1: Los índices cuantitativos (porcentajes de logro, por ejemplo) recogidos por un profesor a partir de los estudiantes de su clase serán, siempre, únicamente descriptivos, lo que significa que *su interpretación debe limitarse a los datos observados en esa población*, sin que se pueda inferir nada para la población en general de la cual se deriva esta muestra NO REPRESENTATIVA (la clase).

Por esta misma razón, a lo largo de este capítulo NO SE HARÁN consideraciones para la generalización de lo que será observado.

[215] El valor 1,96 ("casi dos veces") proviene de una tabla de valores o puntajes Z para una distribución estándar de tipo normal, donde 1,96 es el valor que incluye/excluye al 95%/5% de los valores de la distribución. De esa manera, para un nivel de confianza del 95% (es decir, con un 5% de riesgo de error) se usa el valor 1,96. Para un 99% de confianza se usa el valor de 2.576.

Tabla 7: Límites de confianza de los porcentajes obtenidos a partir de una muestra, según el N de la muestra y autorizando diferentes inferencias hacia el conjunto de la población

Tamaño de la muestra	Cálculo del Error Estándar de Medición de un %			Cálculo	Posibilidades de inferencia hacia el conjunto de la población
Si N =	p*q =	p*q/N =	EEM o Raíz cuadrada de (p*q/N) =	2 * EEM =	Podemos entonces decir, con 95% de probabilidades de no equivocarnos, que la tasa de logro de la población se encuentra entre...
10	0,24	0,024	0,15 (15%)	0,30 o 30%	30% y 90% Anunciar una tasa de incertidumbre así desataría la risa de todos los lectores de nuestro informe.
100	0,24	0,0024	0,05 (5%)	0,10 o 10%	50% y 70% El "rango" de incertidumbre es menor, pero sigue siendo demasiado amplio.
1000	0,24	0,00024	0,015 (1,5%)	0,03 o 3%	57% y 63% (rango = 6%), lo que muestra mayor precisión.
2000	0,24	0,00012	0,01 (1%)	0,02 o 2%	58% y 62% (rango = 4%). Este nivel de incertidumbre es muy aceptable.
4000	0,24	0,00006	0,0077 (0,77%)	0,015 o 1,5%	58,5% y 61,5% (rango= 3%)
8000	0,24	0,00003	0,005 (0,5%)	0,01 o 1%	59% y 61% (rango = 2%). Incluso si duplicáramos el N, la ganancia en precisión sería baja. El rango es considerablemente preciso.

Mala noticia 2: Cuando se quiere comparar, por ejemplo, dos métodos, es común constituir dos grupos de estudiantes seleccionados al azar de la población. Sin embargo, acabamos de ver que aquello está más allá del alcance de un profesor con su clase. Afortunadamente, el método de muestreo representativo no es la única manera de recoger datos relevantes y fructíferos para evaluar el impacto de un *evento pedagógico*. Se pueden formar, por ejemplo, grupos de pares (también llamados "aparejados", "pareados" o incluso "emparejados") en donde cada estudiante del grupo tiene su "equivalente" en el otro grupo (por ejemplo misma edad, mismos requisitos de ingreso, misma motivación, etc.).

Esta operación no es para nada más fácil que el muestreo aleatorio, ya que: a) se debe asegurar que los criterios de coincidencia son los más pertinentes a tener en cuenta (p.ej.: edad, requisitos previos, motivación, etc.); y b) se debe obtener una gran cantidad de información precisa (edad, requisitos previos, motivación, etc.) para cada estudiante. Por otra parte, se debe saber que a pesar de todas estas "precauciones", nunca se podrá garantizar que en cada una de estas "parejas" los dos estudiantes son "equivalentes".

C.5. Investigación en la clase: tres buenas noticias

Buena noticia 1: Un profesor puede hacer estudios longitudinales (en diferentes momentos), en los cuales cada estudiante puede compararse consigo mismo. Ese es el caso del Capítulo 18, en donde los test o pruebas de progreso comparan los cambios observados en 24 pruebas sucesivas aplicadas a los mismos estudiantes a lo largo de 6 años.

Buena noticia 2: El profesor dispone de una gran cantidad de información de "contexto" que le permite interpretar datos (ciertamente, datos limitados a los estudiantes en ese contexto). Sin poseer un "macroscopio" el profesor cuenta con un "microscopio". De esa manera puede superar el problema de la "caja negra" gracias a un enfoque clínico.

Buena noticia 3: El profesor puede realizar estudios contrastados para los mismos estudiantes, lo que le permitirá comprobar si las diferencias *entre estudiantes* son más grandes que las diferencias *intra-estudiantes* (y, por tanto, asociar estas diferencias a los métodos pedagógicos). Es decir, cuando se evalúan diferencias de progreso comparando diferentes momentos del aprendizaje de cada estudiante en el seno de un grupo, y se aprecia que estas son más importantes que las diferencias entre los estudiantes que componen el grupo, es posible atribuir tales diferencias, en algún grado, a las metodologías usadas y sacar conclusiones sobre el aporte de estas.

Caso 3: Las dos instrucciones de aprendizaje vía hipermedia.

El profesor hace la siguiente pregunta (de investigación): "¿Se preparan los estudiantes de manera diferente para los exámenes según el tipo de examen anunciado?". Leclercq y Pierret (1989) eligieron dos ámbitos de contenido, D1 y D2, para aprender a través de un hiper-diaporama. En cada pantalla el estudiante podía solicitar ayudas para el aprendizaje (AA): por un lado, una síntesis (un esquema organizador), y por otro, un cuestionario con Preguntas de Selección Múltiple (PSM) a través del cual podía revisar su comprensión. Los autores dividieron (al azar) una clase de 16 alumnos en dos grupos, G1 y G2, cada uno con 8 alumnos. Al grupo G1 se le dio para aprender D1, diciéndole (sistema oral) que "la evaluación será una presentación oral." A continuación, se le dio a conocer el contenido D2 señalándole que: "El examen constará de Preguntas de Selección Múltiple (PSM) para responder." El grupo G2 recibió las instrucciones opuestas. Por tanto, los dos grupos tenían los mismos dos ámbitos y las mismas dos instrucciones (opuestas).

He aquí las cantidades promedio de utilización espontánea de las ayudas de aprendizaje (AA):

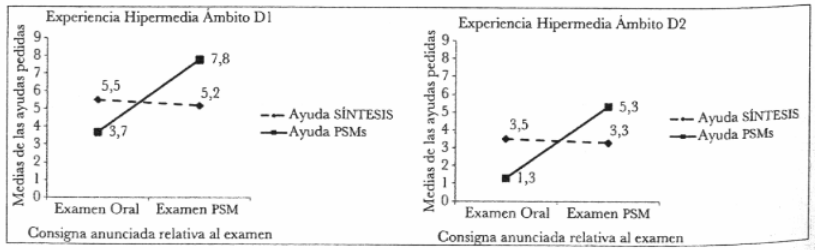

Figura 1: Media de solicitudes espontáneas de ayuda (Leclercq y Pierret, 1989)

Como se ve, con muy pocos estudiantes es posible resaltar un fenómeno de interacción cruzada: el apoyo más buscado (más popular) depende de la instrucción dada en relación a la modalidad de examen. Sin embargo, debemos asegurarnos de que esto no se debe a diferencias en el contenido, y de ahí que recurramos al uso de dos contenidos diferentes presentados a todos. Aunque el contenido de D2 ha atraído a un menor número de solicitudes de ayuda que D1, se observa el mismo fenómeno (interacción cruzada) para ambos ámbitos (ver sección Q).

C.6. Un profesor está solo en su clase... pero no está solo en el mundo

Caso 4: Las respuestas sin (bics azules) y con (bics rojos), y la focalización de la vigilancia.

Responder Preguntas de Selección Múltiple (PSM) con Soluciones Generales Implícitas (SGI) exige mucha más atención de parte de los estudiantes que cuando las PSM son simples (es decir, donde una, entre k soluciones, es la correcta). Por su parte, la vigilancia cognitiva se reparte entre 4 Soluciones Generales Implícitas (SGI): "Otra (o ninguna)", "Todas las anteriores", "Faltan datos (en el enunciado)" y "Absurdo en el enunciado".

Para medir el impacto de la focalización interrogamos, en primer lugar, a los estudiantes pidiéndoles que marcaran sus respuestas con un bolígrafo azul. Luego, en un segundo momento, se les pidió volver a contestar, con un bolígrafo rojo, focalizándose en UNA SOLA de las SGI.

P. Henrotay lo hizo en 2011 en el Ateneo Royale (A.R.) de Spa en Bélgica, con 22 estudiantes de último año de secundaria en una prueba de matemáticas de 20 preguntas PSM + SGI, focalizándose (llegado el momento de los lápices rojos) en el "ABSURDO EN EL ENUNCIADO". Un equipo de profesores de la Universidad Católica de Temuco en Chile (Villagra, Sepúlveda, Torres, Barría, Riquelme, Jara, Sánchez y Leclercq) hizo lo mismo en 2012 con 15 estudiantes universitarios en una prueba de 2° año de psicología, con 12 preguntas PSM + SGI, focalizándose (en el momento de los lápices rojos) en "FALTAN DATOS".

Resultados globales obtenidos en los dos lugares (Leclercq *et al.* 2011; 2013).

Tabla 8: Porcentaje de logro comparado según si la atención es o no focalizada en una solución general implícita (SGI) en una PSM

	UC Temuco Oct. 2012 % de éxito para 15 estudiantes frente a 12 preguntas PSM+SGI (focalización en FALTAN DATOS)		A.R. Spa Nov. 2011 % de éxito de 22 estudiantes frente a 20 preguntas PSM+SGI (focalización en ENUNCIADO ABSURDO)	
	Para las 3 preguntas con "faltan datos"	Para las otras 9 preguntas	Para las 3 preguntas con "enunciado absurdo"	Para las otras 17 preguntas
Lápiz azul (sin focalización)	7%	13%	18%	31%
Lápiz rojo (con focalización)	29%	29%	29%	33%
Ganancia	+22%	+16%	+11%	+2%

A) Las diferencias entre ambos experimentos

El hecho de focalizar la atención y volver a contestar benefició más a los estudiantes universitarios que a los de último año de secundaria. Ambos mejoraron en el caso de

las preguntas cuya respuesta era una SGI, pero los universitarios se vieron además beneficiados con la posibilidad de volver a contestar en el resto de la preguntas. En dichas preguntas (celdas sombreadas), los 15 estudiantes universitarios que rindieron la prueba de 12 preguntas en la UC de Temuco pasaron del 13% al 29% de éxito, es decir, un aumento del 16%. Los 22 alumnos de último año de secundaria que rindieron el test de 20 preguntas en el A.R. Spa pasaron de un 31% a un 33%, es decir, una ganancia de 2%. Esta diferencia puede tener muchas explicaciones: (1) la edad de los estudiantes; (2) el material de la prueba; (3) el número de preguntas; (4) la focalización (no es lo mismo releer centrándose en "faltan datos" que en "absurdo en el enunciado"); (5) las circunstancias (hora del día, fatiga, etc.).

B) LAS SIMILITUDES ENTRE LOS DOS EXPERIMENTOS

Los resultados apuntan en la misma dirección (ambos equipos pueden compartir sus resultados para compararlos: no es necesario que todos repitan la investigación de todos los demás).

Los practicantes pueden identificar clínicamente (ver Capítulo 9) qué otros factores habrían podido intervenir: la dificultad específica de una u otra pregunta o las circunstancias particulares, tales como la preparación de los estudiantes, su nivel de concentración en cada fase (con y sin focalización).

A partir de estos resultados "locales" los profesores pueden formular hipótesis –e imaginar cómo verificarlas– acerca de lo que habrían sido…:

> …los resultados en otra prueba sobre el mismo tema (con otras preguntas) y para los mismos estudiantes, sabiendo que las diferencias podrían deberse a diferencias en la dificultad.
>
> …los resultados de otros estudiantes para la misma prueba (por ejemplo, los estudiantes del año siguiente).

C.7. El diálogo permanente entre la verificación cuantitativa y el enfoque clínico

Hemos visto que a partir de algunas respuestas dadas por grupos reducidos de estudiantes es posible obtener resultados cuantitativos valiosos. El enfoque clínico, por su parte, se centra en casos particulares. Por ejemplo, en el resultado de un estudiante a una pregunta, centrándose no solo en la exactitud de la respuesta sino que, además, en el proceso que conduce a dicha respuesta. El Capítulo 9 (TEM) proporciona una ilustración de este enfoque, donde la información recabada acerca de cada estudiante alcanza una considerable profundidad: la *meta-reflexión* (el autodiagnóstico) del estudiante que lleva a cabo el propio estudiante –"¿por qué estaba tan seguro cuando me equivoqué?" o "¿por qué estaba tan inseguro si mi respuesta era la correcta?"–. Gracias a las notas metacognitivas escritas y al informe retrospectivo, el profesor tiene acceso a una buena parte de tales reflexiones individuales.

C.8. Investigación exploratoria y confirmatoria

La expresión "Verificar una hipótesis" es ambigua. Para algunos (entre quienes nos incluimos) significa "poner la hipótesis a prueba frente a los hechos de la realidad" o "recolectar datos para probarla o comprobar sus fundamentos". Sin embargo, otros dicen: "la hipótesis ha sido verificada" significando con ello que "los datos confirman la hipótesis", y, por desgracia, otorgan a la expresión "confirmar" una connotación "definitiva". Por nuestra parte, creemos que actuamos sobre la base de hipótesis permanentes susceptibles de ser contradichas por la realidad, es decir, ser disminuidas en su fuerza (NO "abandonadas" o "invalidadas"), o bien ser reforzadas (NO "confirmadas").

C.9. Una investigación a nivel de la clase puede ser valorada por un (unos) meta-análisis

Hablamos de una investigación "local" donde a veces usamos el término *indagación* (*inquiry* en inglés), al referirnos a la investigación que el docente lleva a cabo con o a propósito de sus alumnos. Decimos "a nivel de la clase" ya que no necesariamente se realiza "en la clase". Esta investigación local (si es publicada) puede, por lo mismo, ser valorada por un meta-análisis o por varios meta-análisis (ver sección M2). La Tabla 9 y la Figura 2 dan una idea (aunque altamente miniaturizada) de un meta-análisis: presentación resumida (Tabla 9) de las investigaciones seleccionadas (a menudo, más de 20 –aquí son solo 3–), y luego, los valores medios encontrados (Figura 2).

Tabla 9: Ganancias Relativas en los porcentajes de logro gracias a la focalización de la atención (meta-análisis para tres estudios)

	EXPERIENCIAS DE BOLÍGRAFOS AZULES Y DE BOLÍGRAFOS ROJOS. EFECTO (EN %) DE LA RELECTURA	Nº SUJETOS	Nº PREGUNTAS	PRE	POST	GANANCIA	GANANCIA RELATIVA
SIN FOCALIZACIÓN							
Cross y Frary (1977)	PSM simples (4 soluciones, solo una de ellas es correcta)	241	40	25%	33%	8%	*11%*
Henrotay-Math, AR Spa (2011)	PSM con 4 SGI para las 17 preguntas donde la respuesta correcta NO ES "enunciado absurdo"	22	17	31%	33%	2%	*3%*
UC Temuco Psicología 2012	PSM con 4 SGI para las 9 preguntas donde la respuesta correcta NO ES "faltan datos"	15	9	13%	29%	16%	*18%*
CON FOCALIZACIÓN							
Henrotay- Math AR Spa (2011)	PSM con 4 SGI para las 3 preguntas donde la respuesta correcta ES "absurdo"	22	3	18%	29%	21%	*12%*
UC Temuco Psicología 2012	PSM con 4 SGI para las 3 preguntas donde la respuesta correcta ES "faltan datos"	15	3	7%	29%	22%	*24%*

Si ubicamos en un eje estos cinco resultados, vemos que las ganancias relativas son todas positivas. Vemos también que la focalización mejora sensiblemente dichas ganancias relativas.

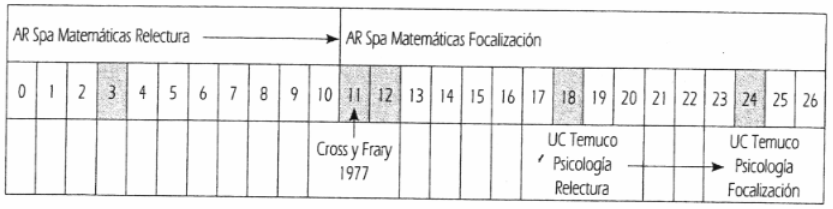

Figura 2: Resumen visual de la Tabla 9

El promedio de Ganancia Relativa sin focalización (relectura) es de 10,6%, mientras que con focalización es de 18%. No obstante, el detalle en la Tabla 9 nos permite apreciar las diferencias que se ocultan bajo una Media de promedios (en este caso: 14,4%) y comenzar a aproximarnos a sus causas.

D. La doble dependencia entre Preguntas y Estudiantes en un test y el modelo de Rasch

D.1. ¿A qué se debe la inestabilidad de los resultados de una pregunta y de un estudiante?

Los resultados frente a una pregunta (o a una prueba) –por ejemplo, la tasa de éxito (o facilidad-dificultad)–, dependen de los estudiantes que respondieron: si son estudiantes de bajo rendimiento, los resultados serán bajos y la dificultad de las preguntas será grande.

Simétricamente, los resultados de un estudiante (o de un grupo o cohorte de estudiantes) dependen de las preguntas que se les formularon: si son preguntas sencillas (por ejemplo, porque el profesor les preparó mucho), sus resultados serán mejores que si se trata de preguntas difíciles. La experiencia pedagógica (o el grado de preparación) de los estudiantes tiene, por lo tanto, un impacto sobre la dificultad de las preguntas, tal cual lo resume la Tabla 10.

Tabla 10: La interdependencia entre la dificultad de los ítems y la capacidad de los estudiantes

	Mayoría de preguntas fáciles	Mayoría de preguntas difíciles
Mayoría de estudiantes de ALTO rendimiento y/o BIEN preparados	Test con resultados altos.	Test con resultados promedio.
Mayoría de estudiantes de BAJO rendimiento y/o MAL preparados	Test con resultados promedio.	Test con resultados bajos.

D.2. La dificultad de una prueba puede ser explicada por múltiples interpretaciones

Debemos tener cuidado con las comparaciones simplistas entre la Media obtenida por los mismos alumnos en diferentes pruebas (por ejemplo, en clases diferentes). Comparaciones del tipo "los estudiantes tienen mejores logros en el curso del profesor A que en el del profesor B". En efecto, podría ser...:

- ... que en el curso impartido por A los conceptos sean más fáciles de aprender que en el curso de B.
- ... que si bien los conceptos del profesor A son muy difíciles (incluso, más difíciles que los de B), el mismo profesor A los ha enseñado tan bien y ha preparado tan bien a sus estudiantes, que estos alcanzan resultados superiores a los del curso de B.
- ... que los conceptos del curso A sean difíciles, pero que A solo haga preguntas sencillas (es decir, que no sean representativas de la verdadera dificultad de los contenidos), dando así lugar a mejores resultados visibles (puesto que no se midieron los aprendizajes difíciles).
- ... que las preguntas de A hayan sido conocidas con anterioridad por los estudiantes (fraudulentamente).
- ... que las diferencias en la dificultad de las preguntas sean el resultado de diferencias en la participación de los estudiantes en los estudios. Por ejemplo, bastaría con que fuera un año de Copa Mundial de Fútbol para que los resultados de ese año sean peores que los del año anterior, debido a que muchos estudiantes pasan muchas horas viendo partidos de Fútbol en lugar de estudiar.

Es muy importante tener en cuenta que siempre nos estamos enfrentando al mismo problema de la comparabilidad y las condiciones en que esta es aceptable.

D.3. El modelo de Rasch y los Wits

El modelo de Rasch (1960) expresa la probabilidad (p) de éxito (p (1)) que posee un estudiante de competencia, capacidad o habilidad *x*, que se enfrentó a una pregunta de Dificultad *y*, lo que se anota como p (1; C=x, D=y). La fórmula es p(1; C=x, D=y) $= e^{C-D} / (1+ e^{C-D})$[216].

Los investigadores han propuesto reemplazar *e* (2,7183) por W (1,24573), un número

que presenta interesantes propiedades numéricas. En efecto, $W^0 = 1$ (obviamente), $W^5=3$, $W^{10}=9$, $W^{-5}= 1/3$ y $W^{-10} = 1/9$. La escala de las habilidades de los estudiantes, así como la magnitud de las dificultades de las preguntas se expresan

[216] "*e*", o número de Euler, es una constante que corresponde a la base de las funciones exponenciales.

en Wits –palabra inspirada en bits (dígitos binarios)– a partir de la letra W [217]. Las competencias de los estudiantes y las dificultades de las preguntas varían entre 0 y 100 Wits (por lo general, entre 30 y 70). En un grupo de preguntas, aquellas que tienen una dificultad "central" muestran un índice cercano a 50 Wits (Leclercq, 1987, p. 37).

$$P\,(1;\,C,D) = \frac{W^{C-D}}{1 + W^{C-D}}$$

Dificultad
Capacidad
(65 - 60)

$$P = \frac{W^{(65-60)}}{1 + W} = 0{,}75$$

Un ejemplo: un estudiante tiene 60 wits (60W) de habilidad o competencia. Si le proponemos una pregunta de dificultad 60W (por lo tanto igual a su capacidad) el estudiante tendrá una probabilidad de 1 sobre 2 (50%) de proporcionar la respuesta correcta (dado que C-D vale 0, entonces $W^{C-D} = 1$: $P\,(1;\,C=60\,;\,D=60) = W^0 / (1 + W^0) = 1 / (1 + 1) = ½$. A partir de aquí se puede calcular fácilmente que, frente a un problema de dificultad…

… igual a su nivel de competencia, un estudiante tiene una probabilidad de éxito del 50%
… de 10 W (wits) superior en dificultad a su competencia, su probabilidad de éxito es del 10%
… de 5 W superior en dificultad a su competencia, su probabilidad de éxito es del 25%
… de 5 W inferior en dificultad a su competencia, su probabilidad de éxito es del 75%
… de 10 W inferior en dificultad a su competencia, su probabilidad de éxito es del 90%

La fórmula se llama "logística" (debido a la expresión $1 / (x + 1)$) y "exponencial" (debido a que la diferencia de C-D es un exponente de W (o de e). Como se muestra en la Figura 3, para cada pregunta las probabilidades se distribuyen en forma de una S mayúscula, con una inflexión de la curva (cambio de dirección o sentido) para el valor de 50%.

Cada pregunta en el modelo de Rasch tiene su Curva Característica del Ítem (Item Characteristic Curve) o CCI, la cual se diferencia de las curvas (CCI) de otras preguntas por el punto de inflexión, es decir, por su posición en la escala de dificultad: cuanto más está a la derecha, la pregunta es más dificil. Constatamos así que todas las preguntas poseen las mismas pendientes. Se trata, de hecho, de una situación ideal donde todas las preguntas tienen un índice de discriminación (r_{pbis}) positivo y alto (véase la sección P).

[217] Para mayores detalles sobre W, que corresponde a la unidad de dificultad de preguntas y de competencia de los individuos, ver Leclercq (1987, pp. 29-41).

El eje horizontal corresponde a la vez a la escala de dificultad de las preguntas y a la de las habilidades o competencias de los estudiantes. Mientras más un estudiante se ubica a la derecha en esta escala horizontal, más elevado es su nivel de competencia y más probable es que tenga éxito en la pregunta indicada.

D.4. La Curva Característica de un Ítem (CCI) de una pregunta

La Figura 3 presenta los CCI (teóricos) para ocho preguntas cuyas dificultades (o puntos de inflexión, en las columnas) varían de 10 a 80 wits. Las habilidades de los estudiantes están en línea (de 10 a 100 wits). Los % de éxito o logro se indican en las celdillas.

	Dificultades de las preguntas (ítems)							
Competencias	10	20	30	40	50	60	70	80
10	0,5	0,100	0,012	0,001	0,000	0,000	0,000	0,000
15	0,750	0,250	0,036	0,004	0,000	0,000	0,000	0,000
20	0,900	0,5	0,100	0,012	0,001	0,000	0,000	0,000
25	0,964	0,750	0,250	0,036	0,004	0,000	0,000	0,000
30	0,988	0,900	0,5	0,100	0,012	0,001	0,000	0,000
35	0,996	0,964	0,750	0,250	0,036	0,004	0,000	0,000
40	0,999	0,988	0,900	0,5	0,100	0,012	0,001	0,000
45	0,100	0,996	0,964	0,750	0,250	0,036	0,004	0,000
50	0,100	0,999	0,988	0,900	0,5	0,100	0,012	0,001
55	0,100	0,100	0,996	0,964	0,750	0,250	0,036	0,004
60	0,100	0,100	0,999	0,988	0,900	0,5	0,100	0,012
65	0,100	0,100	0,100	0,996	0,964	0,750	0,250	0,036
70	0,100	0,100	0,100	0,999	0,988	0,900	0,5	0,100
75	0,100	0,100	0,100	0,100	0,996	0,964	0,750	0,250
80	0,100	0,100	0,100	0,100	0,999	0,988	0,900	0,5
85	0,100	0,100	0,100	0,100	0,100	0,996	0,964	0,750
90	0,100	0,100	0,100	0,100	0,100	0,999	0,988	0,900
95	0,100	0,100	0,100	0,100	0,100	0,100	0,996	0,964
100	0,100	0,100	0,100	0,100	0,100	0,100	0,999	0,988

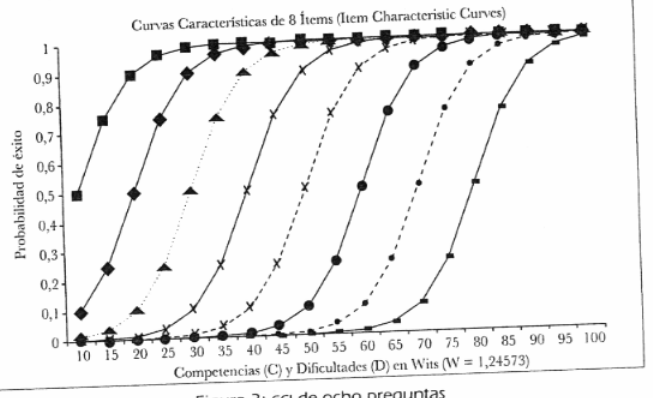

Figura 3: CCI de ocho preguntas

¿Corresponde a la realidad este modelo teórico?

Lord y Novick (1968) lo aplicaron para 2 preguntas (Q1 y Q2) de un examen de matemáticas de 90 preguntas (siendo 90 la puntuación máxima), y a las cuales habían respondido miles de estudiantes.

Ellos dividieron a los estudiantes en grupos: aquellos que tenían una puntuación total de 90, los que tenían una puntuación de 89, etc. (un poco menos de 90 grupos, ya que casi ninguno de los estudiantes tenía menos de 10). A continuación calcularon la tasa de éxito promedio de Q1 y Q2, lo que se muestra en el eje vertical de la Figura 4, para cada grupo (que representa el nivel de competencia de los estudiantes). Cabe señalar que las CCI de estas dos preguntas tienen una curva cuya forma es la logística exponencial. La pregunta Q1 aparece como más fácil que Q2: está situada más a la izquierda, y su punto de inflexión es aproximadamente de 25/90, mientras que el punto de inflexión de Q2 es aproximadamente de 60/90. La curva de Q1 está trunca en la parte inferior: ningún grupo ha logrado Q1 con menos del 35% de respuestas correctas. Los autores hicieron lo mismo con la Q3 de una prueba de lenguaje de 60 preguntas. Observamos que esta curva tiene un piso del 20% (se trata de una PSM de cinco soluciones con una correcta).

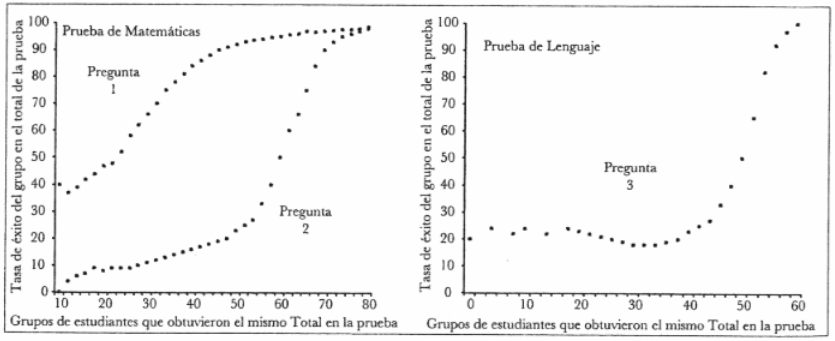

Figura 4: CCI Tres ítems de dos pruebas diferentes

Parte 2. Índices sobre una prueba (un test)

E. Moda, variedad y entropía

Estos índices cuantitativos permiten resumir los resultados de un grupo de estudiantes en una observación (un cuestionario, una prueba, un test, etc.) en CATEGORÍAS NOMINALES.

Caso 5: Atribución en los análisis causales.

Para su actual curso ISE, Leclercq pidió a los estudiantes que le entregaran un informe metacognitivo retrospectivo sobre 3 Tests Espectrales Metacognitivos (TEMS, ver Capítulo 9), donde pidió a los estudiantes intentar explicar las causas de su fracaso en determinadas preguntas. Con una de sus estudiantes (C. Verstegen, 2008), clasificó estas "explicaciones causales" en dos categorías:

(A) la de Rotter (1966): que distingue entre las atribuciones "internas" (YO soy el responsable de este error) y externas (la causa de mi error NO se debe a mí).
(B) la de Leclercq y Poumay (2007): PRE (antes de la prueba), PER (durante la prueba), PRE–PER (ambos).

La combinación de estos dos criterios determina seis categorías nominales que se muestran en la Tabla 11. Ninguna categoría puede considerarse superior a la otra, de modo que si los datos de la Tabla 11 se representan en un gráfico, las cimas de las barras no deben estar conectadas por líneas.

Tabla 11: Distribución (en %) de 166 atribuciones causales de los errores en una prueba, hechas por los estudiantes, organizadas en 6 categorías nominales

	Antes	Antes y Durante	Durante
Causas internas	18,7	3,0	46,4
Causas externas	4,2	1,8	25,9

Las distribuciones de frecuencias en categorías nominales se pueden representar gráficamente en barras horizontales o verticales, como las que se ven en la Figura 5.

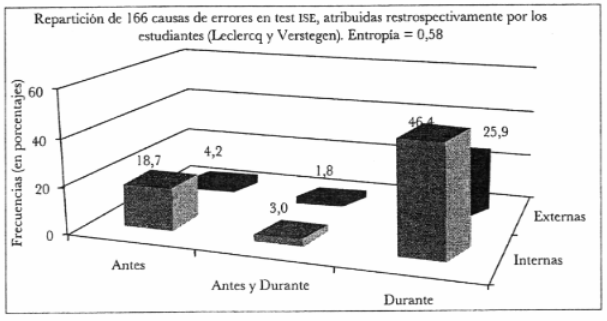

Figura 5: Repartición de seis tipos de explicaciones dadas por los propios estudiantes en relación con las causas de sus errores

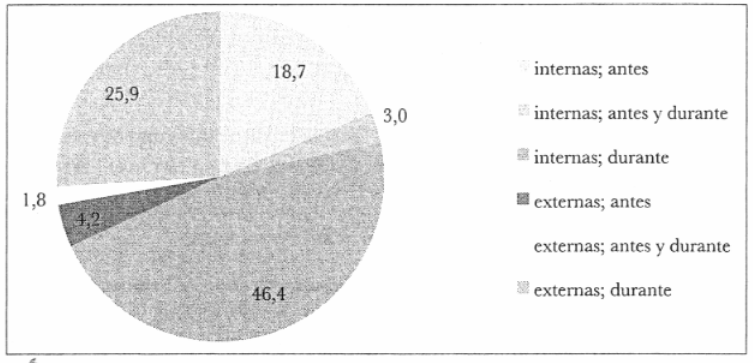

Figura 6: Repartición de seis tipos de explicaciones dadas por los propios estudiantes en relación con las causas de sus errores

La Figura 6 muestra otra forma muy común de representar la distribución de datos nominales: un gráfico de Torta (o de Camembert), recordando así que no existe un orden entre las categorías. No obstante, en este ejemplo (datos de Tabla 11), las barras resultan más elocuentes o ilustrativas de acuerdo con la estructura de los datos.

E.1. La Moda

La Moda es el valor con mayor frecuencia en una distribución de datos. En nuestro ejemplo, la frecuencia 46,4% se observó en la categoría "Durante (PER) / Interna". También podemos decir que "La categoría modal es PER / Interna".

E.2. La Variedad (o diversidad)

La variedad es la cobertura de la distribución de los datos en las categorías nominales. En las figuras 5 y 6, así como en la Tabla 11, se ve que las seis categorías están representadas en la distribución; todas fueron observadas, es decir, ninguna tiene valor 0.

E.3. La entropía y neguentropía

La entropía y neguentropía son índices inversos de dispersión y de concentración de categorías nominales. En el ejemplo se observa una concentración en las explicaciones causales PER (durante), y podría ser deseable que el análisis causal que realizan los estudiantes de sus errores se refiriera más a una causa anterior (PRE), cuando el profesor no está a su lado para ayudarles. Si, gracias a alguna intervención pedagógica como el Test Espectral Metacognitivo (TEM), ocurriera una mayor atribución a causas PRE, la entropía (igualdad) aumentaría.

Tabla 12: Términos posibles (sinónimos) para designar la entropía y la neguentropía, pero en contextos y con connotaciones diferentes

Neguentropía o desigualdad (u orden o estructura o información o concentración) $= \Sigma p \log(p)$	*Entropía* o igualdad o *–Neguentropía* (o desorden o desorganización o dispersión o *descentración*) $= -\Sigma p \log(p)$
Esta suma es siempre negativa, ubicada entre 0 (concentración máxima) y –1 (dispersión máxima)	*Este índice es siempre positivo*, y se ubica entre 0 (concentración máxima) y 1 (igualdad máxima)

Este índice (entropía) es poco utilizado en pedagogía... y sin embargo merecería serlo, tal como trataremos de mostrar en los dos casos siguientes. La razón de este poco uso se debe probablemente a los términos que, por desgracia, se han utilizado hasta el momento para hablar de este índice. La noción de entropía proviene de la termodinámica, en donde se observó que un gas tiende a dispersarse y ocupar todo el espacio disponible, es decir, a propagarse (Claussius, Sadi Carnot). De ahí el término "desorden, dispersión, incoherencia" en termodinámica. Sin embargo, al aplicar la fórmula de la entropía (y el concepto) en las ciencias humanas, se mantuvieron estos mismos términos con una connotación negativa. Ahora bien, bastaría con utilizar otras palabras que pueden operacionalizar el concepto de entropía cuando este se aplica a otras situaciones (humanas en este caso), tal como se muestra en la Tabla 13.

Tabla 13: Variante de la Tabla 12

	En termodinámica	En educación
Entropía	Desorden, Desorganización o Dispersión	Repartición, Co–responsabilidad, Descentración, Desenfoque
Neguentropía o Negantropía	Orden, Estructura, Información	Concentración, Monopolización, Reducción (de libertad), Restricción

"Nombrar mal las cosas es agregar desgracia en el mundo"
(Albert Camus, 1913-1960)

Caso 6: Tomar la palabra.

En su curso de inglés para estudiantes cuya lengua materna es el español, el profesor organiza las "mesas de conversación en inglés", con 10 participantes cada una y una duración de 2 horas. Pide que cada uno de los 10 participantes tome más o menos el mismo número de veces la palabra, para tener el mismo número de oportunidades que los otros de hablar y practicar el inglés.

Las dos preguntas que aparecen son: "(1) ¿Hay estudiantes (o categorías nominales) que toman más seguido la palabra que otros?" y "(2) ¿Cuál es la distribución o igualdad/desigualdad en el uso de la palabra?". Imaginemos cuatro sesiones sucesivas (A, B, C y E) con sus mesas de conversación de 10 estudiantes (E1, E2, E3, etc.), y pensemos que en cada una de ellas hubo un total de 40 usos de la palabra en público cada vez. La Figura 7 muestra la distribución de este uso de la palabra.

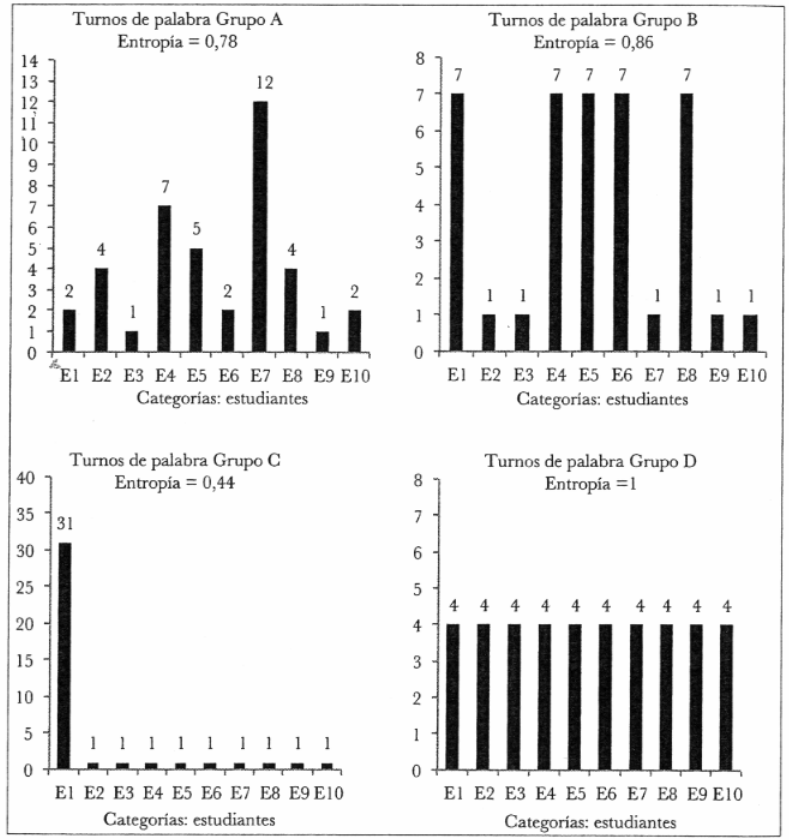

Figura 7: Reparticiones (ficticias) de uso de la palabra en cuatro sesiones (o grupos)

En la sesión (grupo)...

A, la categoría modal es E7 (frecuencia = 12) y la entropía 0,78.
B, la categoría modal es E1, E4, E5, E6, E8 (frecuencia = 7) y la entropía 0,86.
C, la categoría modal es E1 (frecuencia = 31) y la entropía 0,44.
E, la categoría modal es compartida por todos (frecuencia = 4) y la entropía 1.

Con humor, el profesor de inglés dice: "Cuando yo monopolizo la palabra, me constituyo en un grupo "D" (con D de "Dominación"), y me otorgo 391 de 400 palabras dadas, dejando a cada uno de los otros miembros del grupo una sola oportunidad para hablar, lo cual hace que la entropía (igualdad) sea de 0,06, es decir, casi cero". Aquí, el problema es la igualdad, un tema poco considerado en educación, donde a menudo

se promueve solo la competencia (el orden, la detección de los líderes) y las diferencias en los resultados entre estudiantes (el valor discriminativo de los resultados). El caso 6 ilustra las diferentes dinámicas de una sesión a otra, lo que tendrá un impacto diferencial en los resultados.

F. La Mediana y el rango semi-intercuartil de una distribución ORDINAL

Caso 7: Creencias (*Opiniones*) *epistemológicas*. Consideremos un cuestionario que contiene preguntas de desarrollo epistemológico (Perry, 1985), del tipo: *El genio es 10% de talento y 90% de trabajo duro.*
El 5 de octubre de 2009, 198 estudiantes del curso *ISE* contestaron, al inicio de aquel (*PRE*), 10 preguntas de este tipo. Luego, el 14 de diciembre de 2009, al final del curso (*POST*) contestaron las mismas preguntas. En ambos momentos debieron elegir respuestas entre 0 ("Desacuerdo total") y 9 ("Acuerdo total").

F.1. Frecuencias acumuladas

Tabla 14: Frecuencias (líneas de arriba) y frecuencias acumuladas (líneas de abajo en cada caso) para dos distribuciones de respuestas de los mismos estudiantes a la misma pregunta (PRE y POST)

	0 Desacuerdo Total	1 Fuerte desacuerdo	2 Desacuerdo	3 Más bien desacuerdo	4 Ligero desacuerdo	5 Ligero acuerdo	6 Más bien acuerdo	7 Acuerdo	8 Fuerte acuerdo	9 Acuerdo Total
PRE	0	30	32	42	29	19	15	17	11	3
	0	30	62	104	133	152	167	184	195	198
POST	1	17	26	29	23	13	36	29	18	6
	1	18	44	73	96	109	145	174	192	198

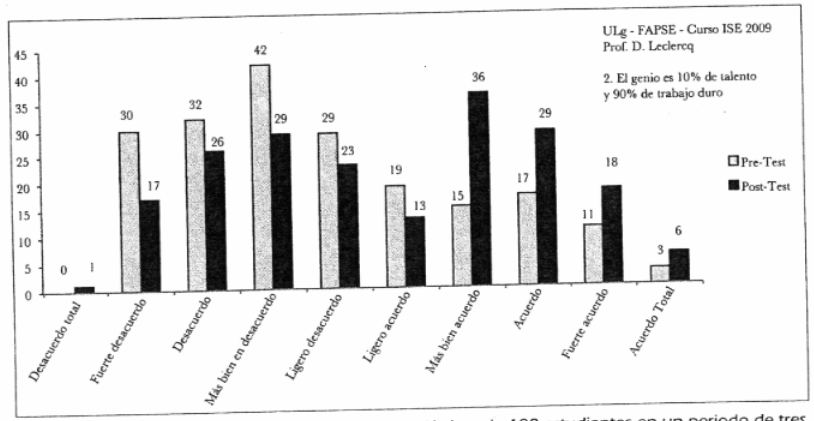

Figura 8: Evolución (PRE-POST) de creencias epistemológicas de 198 estudiantes en un periodo de tres meses a partir de un curso impartido en 2009

La hipótesis de esta investigación-acción era: "Los estudiantes evolucionan en la dirección de 'Acuerdo total' después del curso". Adicionalmente, para saber si dicha evolución es "causada por el curso", se les preguntó también "por qué" (tales resultados no se analizan aquí). La Figura 9 presenta los polígonos de frecuencia de estas dos distribuciones de datos de una escala ordinal:

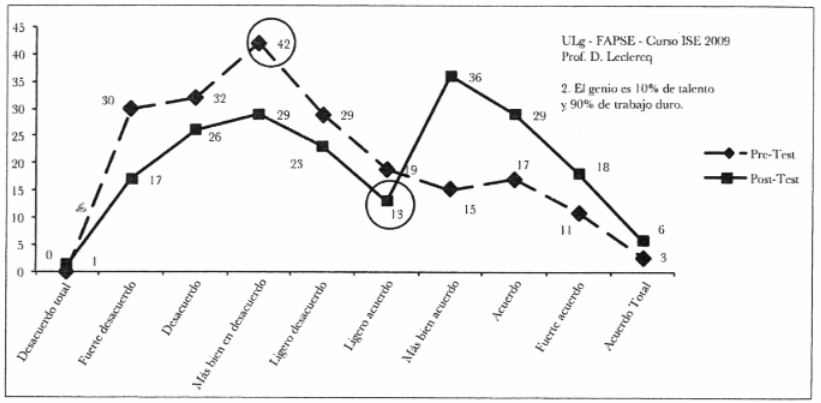

Figura 9: Evolución de las distribución PRE y POST a una de las preguntas epistemológicas para 198 estudiantes en 2009. Las Medianas han sido rodeadas por un círculo

F.2. La Mediana

Las dos *Modas* (o puntos sobresalientes) son fáciles de identificar. Sin embargo, las *Medianas* son menos "visibles": para identificarlas se debe contar (en la Tabla 14). Por lo mismo, las categorías medianas están encerradas en un círculo (Figura 9). En la Tabla 14 hay 198 resultados, donde el percentil 10 es el valor (o la categoría o el puntaje) obtenido por el estudiante que ocupa el rango 19,8 en la distribución (redondeado a 20). En el PRE test dicho valor corresponde a la categoría 1 y en el POST a la categoría 2. De la misma manera, el percentil 90 es la categoría del estudiante que ocupa el rango 179,2 en la distribución. Se aprecia que en el PRE test dicha categoría corresponde a 7 y en el POST es 8. Por supuesto, *el percentil 50* (es decir *la Mediana*) se ubica en el rango 99 (198/2). Para el PRE test, la Mediana es la categoría 3 ("más bien en desacuerdo") mientras que para el POST test lo es la categoría 5 ("ligero acuerdo").

F.3. La distancia intercuartiles

Este índice permite cuantificar la dispersión de las frecuencias para categorías ordinales. Se requiere clasificar las respuestas (poniéndolas en orden), desde la más "baja" (o negativa) a la más alta, y luego contar. Así, vemos qué categoría nominal…

- ocupa el rango 25 de 100 (en nuestro ejemplo, la 44,5ava de 198) y nombramos esta categoría como Q1 (o primer cuartil)

- ocupa el rango número 50 de 100 (la 99ava de 198) y nombramos esta categoría como Q2 (el segundo cuartil) o la mediana
- ocupa el rango 75 de 100 (la 143,5ava de 198) y nombramos esta categoría como Q3 (el tercer cuartil)

La diferencia intercuartil se calcula entonces entre Q3 y Q1. En este caso, dicha distancia corresponde a 3 categorías. La desviación semi-intercuartil es de 3/2 = 1,5 categorías en PRE test y 4/2 = 2 categorías POST test.

Tabla 15: Cómo calcular los Cuartiles Q1, Q2 (la mediana) y Q3

	0 Desacuerdo total	1 Fuerte desacuerdo	2 Desacuerdo	3 Más bien desacuerdo	4 Ligero desacuerdo	5 Ligero acuerdo	6 Más bien acuerdo	7 Acuerdo	8 Fuerte acuerdo	9 Acuerdo total
PRE	0	30	32	42	29	19	15	17	11	3
Acumulado	0	30	62	104	133	152	167	184	195	198
	La categoría ocupada por el 44ave y 45ave estudiante sobre 198 o Q1 es En PRE: 2. Desacuerdo En POST: 2. Desacuerdo			La categoría ocupada por el 99ave estudiante sobre 198 o Q2 es En PRE: 3. Más bien desacuerdo En POST: 4. Ligero desacuerdo			La categoría ocupada por el 143ave y 144ave estudiante sobre 198 o Q3 es En PRE: 5. Ligero acuerdo En POST: 6. Más bien acuerdo			
POST	1	17	26	29	23	13	36	29	18	6
Acumulado	1	18	44	73	96	109	145	174	192	198
	0 Desacuerdo total	1 Fuerte desacuerdo	2 Desacuerdo	3 Más bien desacuerdo	4 Ligero desacuerdo	5 Ligero acuerdo	6 Más bien acuerdo	7 Acuerdo	8 Fuerte acuerdo	9 Acuerdo total

G. La Media (M) o promedio, la Desviación Estándar (DE), histogramas y asimetría de una distribución MÉTRICA

G.1. Los datos de base

La Tabla 16 presenta:

a) en su cuadro central, el valor *binario* (1 y 0) para cada respuesta dada por cada uno de los 13 estudiantes que respondieron las 10 preguntas de una prueba, 1 significando "éxito" y 0 significando "error u omisión".
b) en la *fila NEp*, el número total de éxitos (máximo= 13) para cada una de las 10 preguntas. Al final se presenta la Media (en este ejemplo 7,7). En la *fila TEp* se presentan los *porcentajes de Éxito* por pregunta, con su respectiva Media en el total (aquí 0,59).

c) en la *columna NEs* se presenta, para cada uno de los 13 estudiantes o sujetos, el número de éxitos en la prueba (máximo=10) con su respectiva Media por sujetos (aquí 5,92). En la *columna TEs* se entregan los *Porcentajes de Éxito* con su Media (aquí 0,59).

Tabla 16: Datos de base de un test (prueba) de demostración. NS= n° de sujetos. NP= n° de preguntas

NS=	13		PREGUNTAS											
NP=	10		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		
	NE	NEp	7	9	5	11	7	10	4	9	7	8	7,70	Media
Estudiantes	NEs	TE	0,54	0,69	0,38	0,85	0,54	0,77	0,31	0,69	0,54	0,62	TEp	0,59
E1	4	0,4	1	0	1	1	0	1	0	0	0	0		
E2	9	0,9	1	1	1	1	1	1	1	1	0	1		
E3	6	0,6	0	1	0	1	1	1	1	0	0	1		
E4	6	0,6	1	1	1	0	1	1	0	1	0	0		
E5	3	0,3	0	0	0	1	0	0	0	1	0	1		
E6	5	0,5	0	1	1	1	0	0	1	0	1	0		
E7	7	0,7	0	1	0	1	1	1	0	1	1	1		
E8	5	0,5	1	1	0	0	0	1	0	1	0	1		
E9	5	0,5	0	0	0	1	0	1	0	1	1	1		
E10	2	0,2	0	0	0	1	0	0	0	0	1	0		
E11	7	0,7	1	1	0	1	1	1	0	1	1	0		
E12	10	1	1	1	1	1	1	1	1	1	1	1		
E13	8	0,8	1	1	0	1	1	1	0	1	1	1		
		TEs												
MEDIA	5,92	0,59												
Desv. Est.	2,20													

G.2. El histograma y el polígono de frecuencias de la notas de los estudiantes

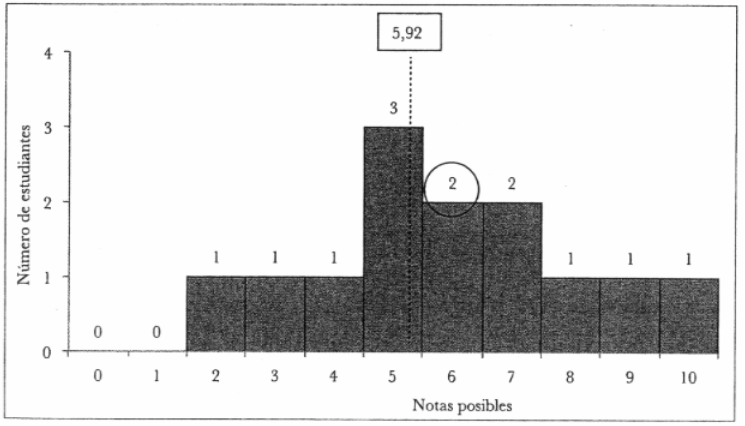

Figura 10: Histograma de la distribución de las notas de los estudiantes

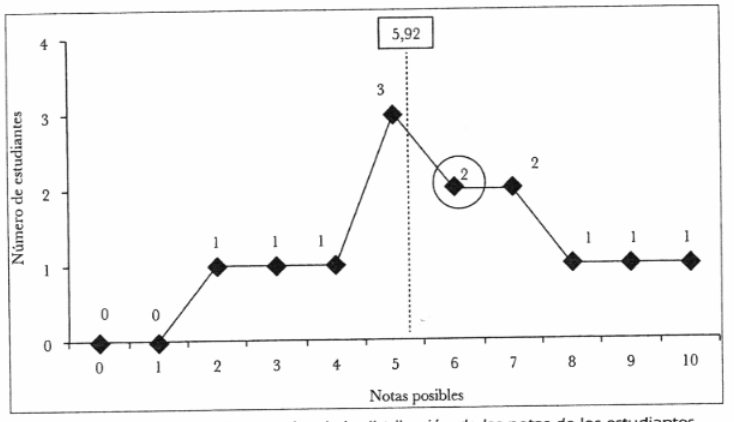

Figura 11: Polígono de frecuencias de la distribución de las notas de los estudiantes

La Figura 10 es un *histograma* mientras que la Figura 11 se llama *polígono de frecuencias*, y ambas grafican los datos de la Tabla 16. En el caso del polígono, una línea ininterrumpida pasa por las cimas de los rectángulos del histograma. Este tipo de presentación es menos preciso puesto que los lectores podrían pensar que hay valores intermedios entre dos puntos de la línea, lo que no es el caso. Sin embargo, el *polígono de frecuencias* se utiliza cuando varias distribuciones están superpuestas, lo que ayuda al lector a identificar y comparar, rápidamente, varias distribuciones de manera simultánea. Las barras (o bastones) del *histograma* aparecen juntas porque el total (100%) de las observaciones (aquí 13) corresponden a la suma de las *áreas cubiertas* por el histograma y al área bajo la curva constituida por el *polígono de frecuencias*.

G.3. La Media (o promedio): el principal Índice de tendencia central

Tabla 17: Definición de los índices de tendencia central y sus valores en el ejemplo

Índices de tendencia central	En el ejemplo (figuras 10 y 11), corresponden a...
La nota modal o "la Moda" es la nota más frecuente.	En las figuras 10 y 11 aparece claramente que la *Moda es la nota 5.*
La *nota Mediana* es la nota obtenida por el sujeto (estudiante) que ocupa la posición "central" en las notas ordenadas. Equivale a la nota que divide la distribución en dos partes iguales.	Con 13 estudiantes en este caso, la mediana corresponde al estudiante que tiene el orden 7. Se ve que este estudiante tiene una nota 6, la que aparece indicada con un círculo. *La Mediana equivale aquí a la nota 6.*
La Media de las notas (o promedio) se calcula por la fórmula $\Sigma x / NS$ donde Σx es la suma de todos los resultados y NS el número de sujetos, o estudiantes en este caso.	En este caso la *Media es 5,92*, y no corresponde a una de las 11 notas posibles (de 0 a 10), sino que se trata de un *índice calculado*. Su ubicación en la distribución de las notas aparece indicada con una línea punteada.