UN EJEMPLO PRÁCTICO SOBRE LA DISCORDANCIA ENTRE LA SIGNIFICACIÓN ESTADÍSTICA Y LA SIGNIFICACIÓN SUSTANTIVA CON RELACIÓN A LA DECISIÓN DE LA HIPÓTESIS NULA

Clemente Rodríguez Sabiote
Oswaldo Lorenzo Quiles
Juan José Navarro Hernández

Facultad de Educación y Humanidades. Campus de Melilla
Universidad de Granada

RESUMEN

El presente trabajo aboga por el uso de estrategias alternativas a la significación estadística que formarían parte de lo que se ha venido en denominar la significación sustantiva. Por otra parte se ha determinado, mediante un ejemplo práctico, la escasa concordancia que se produce entre la SS y la SE respecto a la comisión del error tipo II, es decir, la aceptación de la hipótesis nula cuando ésta en realidad es falsa.

ABSTRACT

This work plads for the use of alternative strategies to the stadistical significance that would of what has been come in denominating the substantive significance. On the other hand, it has been determined, by means of a practical example, the scarce agreement that place between the stadistical significance and the sustantive, regarding the commision of the error type II, that is to say, the acceptance of the null hypothesis when, in fact, is false.

1. INTRODUCCIÓN

La llegada del método científico a la disciplina educativa supuso una cierta liberación de ésta respecto al epígono filosófico-especulativo al que hasta entonces estuvo unida. Los procedimientos experimentales entraron en el terreno pedagógico en el último cuarto del siglo XIX y hubieron de hacerlo en liza con una fuerte oposición contra su uso e implantación en el terreno educativo (García Hoz, 1994).

Sin embargo, el método experimental consigue ir abriéndose camino progresivamente en la disciplina pedagógica, naciendo así la investigación educativa. Para Bartolomé (1984), se dan cuatro aspectos básicos que coadyuvan en la génesis de la Pedagogía Experimental:

1. *"La preocupación, presente desde antiguo, aunque aparezca con mayor énfasis a partir de Rousseau y Pestalozzi, de asentar la educación sobre bases empíricas y no meramente especulativas.*
2. *La introducción progresiva del método experimental en ciencias afines a la educación.*
3. *La vinculación inicial con la psicología.*
4. *La influencia primigenia de la paidología".*

Una vez en marcha la investigación educativa alcanza cotas de crecimiento inimaginables. La fructífera colaboración de la estadística, la incorporación de las nociones de test y escalas... contribuyen sin duda a este espectacular desarrollo. Pero en este proceso se produce un hecho que a partir de la década de los sesenta genera una verdadera conmoción intelectual y social: el cuestionamiento del reduccionismo positivista. Efectivamente, empiezan a levantarse voces en contra del dogma de que sólo las realidades medibles constituyen la base adecuada de la ciencia. El peligroso traspaso, sin adecuación previa, del método de indagación usado en las ciencias "duras" a las sociales determina tres consecuencias importantes:

1. La sobrevaloración del método experimental en la investigación educativa, hasta el punto de que se identificaba investigación educativa con pedagogía experimental.
2. La necesidad de admitir otros saberes, además de los puramente numérico-cuantitativos.
3. La posibilidad de estudiar los aspectos no puramente cognitivos de la personalidad humana.

Estas dos últimas consecuencias determinan la aceptación de otras vías para llegar al conocimiento en la investigación educativa, además de la meramente cuantitativa. Frente al principio de distinción o exclusividad paradigmática, hay quienes defienden la complementariedad (Shulman, 1985 ; Walkers y Evers, 1990) y quienes incluso niegan la validez del mismo concepto de paradigma (Teoría Criticista). Otros, más recientemente, han acuñado el término *"inconmensurabilidad"* (De la Orden y Mafokozi, 1997) para referirse a una postura contraria a las rigideces y exclusivismos metodológicos; y hay quienes finalmente proponen nuevos retos metodológicos para el nuevo siglo XXI que se avecina como respuesta a las demandas de la sociedad del futuro (Dendaluze, 1998).

En este contexto de cuestionamiento generalizado debemos incardinar lo que Fernández Cano (1995) denomina *"la falacia de la significación estadística"* (de ahora en adelante SE) de los hallazgos, es decir, la tendencia a aceptar/rechazar la hipótesis nula (Ho) tomando como base un nivel de probabilidad "*p*". El dilema es bien antiguo, ya Bernoulli, a principios del S. XVIII, se cuestionó la adecuabilidad del contraste de significación estadística y su uso en la investigación científica (Hacking, 1965 citado por Fernández Cano, 1995). En muchos otros trabajos (Asher, 1993; Bakan, 1967; Chow, 1988; Cronbach, 1975; Guttman, 1985; Harcum, 1989; Huberty, 1987; Meehl, 1967; Morrison y Henkel, 1970; Signorelli, 1974; Skinner, 1956 y Thompson, 1989) igualmente se despliegan una colección de argumentos para, cuando menos, tomar con cautela los resultados a los que podemos llegar a través de la significación estadística. Desde el punto de vista lógico, y frente a la generalizada opinión en contra, la significación estadística no es una condición necesaria, ni suficiente, de la significación educativa de los resultados. Es más, *"en muchos casos en los que se aplica la estadística inferencial la decisión sobre significación estadística no es suficiente, porque lo que interesa no es tanto si hay algún cambio, independientemente de su tamaño, sino si el cambio es suficiente. No tener presente el alcance real de una conclusión derivada de la significación estadística ha llevado a ampliar erróneamente el campo semántico del término, incluyendo el concepto de relevancia de una diferencia"* (León, 1984 ; Oakes, 1986 citados por Botella y Barriopedro, 1995).

Esta afirmación no es un argumento frívolo ni gratuito en contra de la intocable frontera del $p \leq 0.05$ (significación socialmente aceptada). A partir de este valor todo toma sentido con la significación estadística, como si por ejemplo dos medias aritméticas de 5.15 y 5.18 pertenecientes a dos grupos de alumnos (niños y niñas) que

han obtenido un $p \leq 0.05$ en desempeño matemático, a través de los numerosos contrastes de hipótesis existentes, fuesen sustantivamente distintos. Puestos en la realidad educativa, 0.03 centésimas de punto ¿es realmente una diferencia lo suficientemente sustantiva como para indicar que los niños y niñas tienen un desempeño matemático diferente?. Esta pregunta se refiere a la relevancia de un efecto, y es bien distinta a la que subyace en los contrastes de significación. Por supuesto que el concepto de relevancia o SS incluye al de SE. Por tanto, la respuesta a la cuestión por nuestra parte es que al menos deben admitirse con cautela las afirmaciones que (tomando como ejemplo lo anterior) se hagan acerca de las diferencias entre niños y niñas en el desempeño matemático. Para conocer más y mejor sobre las razones por las que la SE es un procedimiento cuestionable se puede consultar Fernández Cano (1995).

Llegado este momento, el lector podría preguntarse de qué forma un investigador puede establecer si los resultados de un estudio son consistentes o no con la predicción, es decir, si las diferencias entre las muestras son debidas, o no (inazarosa) al azar, o sea, si acepto Ho (hipótesis nula) o bien H_1 (hipótesis alternativa).

Las posibilidades son variadas, pero entre todas ellas destaca una idea que debe conducirlas: el sentido común. También aquí aconsejamos consultar la obra de Fernández Cano (1995) para un examen más profundo de esta cuestión.

Por nuestra parte vamos a intentar mostrar al lector, a través de un ejemplo real, tres aspectos que nos parecen relevantes:

1. La utilización de algún tipo de estrategia no basada en la probabilidad, para denotar diferencias significativas entre grupos.
2. La discordancia existente entre la SE y SS respecto a la aceptación/rechazo de la hipótesis nula.
3. La comisión del error tipo II o ß, es decir, afirmar que no hay diferencias significativas cuando en realidad sí las hay, tomando como criterio de verdad/falsedad la SS.

2. PROPUESTA PRÁCTICA PARA CONTRASTAR LAS DIFERENTES DECISIONES QUE SE PUEDEN TOMAR RESPECTO A LA HIPÓTESIS NULA (H_o) A PARTIR DE LAS APROXIMACIONES DE SE Y SS

2.1. Contextualización

El ejemplo que mostramos a continuación corresponde a un conjunto de resultados extraídos de la tesis doctoral de uno de los autores del presente trabajo[1]. En él puede apreciarse con clara nitidez las excelsas posibilidades de la triangulación metodológica como estrategia de recogida y validación de información. En la tabla que presentamos más adelante apreciaremos la opinión de cuatro agentes distintos (alumnos, profesores, miembros del PAS – Personal de Administración y Servicios – y equipo decanal), recogida mediante dos instrumentos diferentes (relatos narrativos y entrevistas en profundidad), sobre un tema determinado: los desajustes y disfunciones asociados en mayor o menor medida con la implantación de los nuevos planes de estudio en la Facultad de Ciencias de la Educación de la Universidad de Granada. Podemos observar cómo estos desajustes se articulan en torno a seis categorías generales de problemas. Lo interesante, precisamente de la doble triangulación ha sido comparar agentes por categoría general en diadas y observar los puntos en común y las divergencias sobre dichas categorías.

2.2. Propuesta

Estos son los resultados obtenidos en la doble triangulación. A cada comparación por binomio de agentes le han sido aplicada las dos aproximaciones, más exactamente a las proporciones comunes (p) y no comunes (q). Para la SE por la naturaleza de los datos objeto del análisis con presencia de frecuencias esperadas menores de 5 (fe<5) en más de un 20% de los casos hemos decidido utilizar una prueba alternativa al X^2, como es el test Binomial, so pena de desvirtuar los resultados obtenidos (Pick y López, 1994; Siegel, 1991 ; Seoane y otros, 1987). Como alternativa complementaria a la SE, es decir, como SS, hemos asumido las diferencias entre problemas comunes (p) y no comunes (q) en términos de tamaño del efecto, de manera que hemos considerado significativamente sustantivas aquellas comparaciones con una proporción común $p \leq 30$, o sea, aquellas con un 70% o más de porcentaje no común (q*100). Estadísticamente podría decirse que la hipótesis nula sería:

1 Tesis doctoral en curso de C. Rodríguez Sabiote.

Ho: μ (p) ≥ 0.30 o también Ho: μ (q) ≤ 0.70, frente a la lógica del contraste significativo donde la hipótesis nula (Ho) se plantearía como Ho: μ (p) = 0.5 o también Ho: μ(p) =μ(q).

Esta alternativa podría formar parte de lo que con anterioridad hemos denominado de *sentido común* y desde luego, según nuestra opinión, se aproxima más a la realidad educativa. Estos son los resultados obtenidos en las múltiples comparaciones:

Categoría General	Comparaciones de agentes posibles	Nº de subcategorías comunes a los agentes comparados	Nº de subcategorías no comunes a los agentes comparados	Proporción común (p)	Proporción no común (q)	Significación estadística
PROFESOR	1) Alumnos Vs Eq. Decanal	1	9	**0.10****	0.90	**0.039***
	2) Profesores Vs Eq. Decanal	1	9	**0.10****	0.90	**0.021***
	3) Alumnos Vs Profesores	4	9	0.31	0.69	0.267
GESTIONES ADMINISTRATIVAS	4) Profesores Vs Eq Decanal	1	3	**0.25****	0.75	0625
	5) Profesores Vs PAS	1	3	**0.25****	0.75	0.625
	6) Profesores Vs Alumnos	1	7	**0.13****	0.89	0.070
	7) Alumnos Vs PAS	2	8	**0.20****	0.80	0.190
ALUMNO	8) Profesores Vs Eq. Decanal	1	6	**0.14****	0.86	0.125
	9) Alumnos Vs Eq. Decanal	2	7	**0.22****	0.78	0.180
	10) Alumnos Vs PAS	1	8	**0.11****	0.89	**0.039***
	11) Alumnos Vs Profesores	4	7	0.36	0.64	0.549
ASIGNATURA	12) Alumnos Vs Eq. Decanal	2	12	**0.14****	0.86	**0.013***
	13) Profesores Vs Eq. Decanal	2	6	**0.25****	0.75	0.289
	14) Alumnos Vs Profesores	3	10	**0.23****	0.77	0.092
	15) Alumnos Vs PAS	2	9	**0.18****	0.82	0.065
HORARIO	16) Alumnos Vs PAS	1	4	**0.20****	0.80	0.375
	17) Profesores Vs Eq. Decanal	1	1	0.50	0.50	1.000
	18) Alumnos Vs Profesores	1	4	**0.20****	0.80	0.375
	19) Alumos Vs Eq. Decanal	2	3	0.40	0.60	1.000
PERSONAL DE ADMINISTRACIÓN	20) Alumnos Vs PAS	1	2	0.33	0.67	1.000

Tabla 1. Resultados obtenidos a partir de la tabla de comparaciones multilaterales y utilizando para el contraste la prueba binomial y un criterio de significación sustantiva. (*Diferencias significativas con un valor de significación de 0.05. **Diferencias sustantivamente significativas: aquellas con una proporción común menor o igual a 0.30)

A la luz de los resultados obtenidos en la anterior tabla, queremos llamar la atención sobre, por ejemplo, la comparación 6: Profesores Vs Alumnos: 1 problema común y 7 no comunes arrojan un p= 0.07, tras el contraste con una prueba no paramétrica como el test binomial. Estadísticamente podemos afirmar que los agentes comparados perciben de igual forma la categoría general a que se refiere la comparación, ya que aceptamos la Ho. Sustantivamente creemos que sólo un problema común indica más bien lo contrario, es decir, una percepción diferente sobre la categoría en cuestión. Lo mismo podríamos decir de las comparaciones 7, 8, 9, 13...

Como podemos comprobar, la utilización de una u otra aproximación cambia sustancialmente las decisiones sobre aceptación/rechazo de la Ho, y en definitiva sobre la afirmación de que la percepción de los agentes es más bien distinta que parecida. En la siguiente tabla mostramos la decisión sobre la aceptación de la Ho en cada comparación, según la aproximación contemplada:

Categoría General	Comparaciones de agentes posibles	SE	SS
PROFESOR	1) Alumnos Vs Eq. Decanal	RECHAZAR	RECHAZAR
	2) Profesores Vs Eq. Decanal	RECHAZAR	RECHAZAR
	3) Alumnos Vs Profesores	ACEPTAR	ACEPTAR
GESTIONES ADMINISTRATIVAS	4) Profesores Vs Eq Decanal	ACEPTAR	RECHAZAR
	5) Profesores Vs PAS	ACEPTAR	RECHAZAR
	6) Profesores Vs Alumnos	ACEPTAR	RECHAZAR
	7) Alumnos Vs PAS	ACEPTAR	RECHAZAR
ALUMNO	8) Profesores Vs Eq. Decanal	ACEPTAR	RECHAZAR
	9) Alumnos Vs Eq. Decanal	ACEPTAR	RECHAZAR
	10) Alumnos Vs PAS	RECHAZAR	RECHAZAR
	11) Alumnos Vs Profesores	ACEPTAR	ACEPTAR
ASIGNATURA	12) Alumnos Vs Eq. Decanal	RECHAZAR	RECHAZAR
	13) Profesores Vs Eq. Decanal	ACEPTAR	RECHAZAR
	14) Alumnos Vs Profesores	ACEPTAR	RECHAZAR
	15) Alumnos Vs PAS	ACEPTAR	RECHAZAR
HORARIO	16) Alumnos Vs PAS	ACEPTAR	RECHAZAR
	17) Profesores Vs Eq. Decanal	ACEPTAR	ACEPTAR
	18) Alumnos Vs Profesores	ACEPTAR	RECHAZAR
	19) Alumos Vs Eq. Decanal	ACEPTAR	ACEPTAR
PERSONAL DE ADMINISTRACIÓN	20) Alumnos Vs PAS	ACEPTAR	ACEPTAR

Tabla 2. Decisión sobre la Ho según la aproximación de SE/SS

Parece un hecho fehacientemente comprobado que la utilización de una u otra vía genera conclusiones distintas. En este caso, sin pretender una estigmatización de la SE, creemos que es más útil y realista el uso de la SS, pues hemos obtenido el mayor partido a las estrategias de triangulación.

Pero este trabajo quedaría incompleto si no fuésemos capaces de cuantificar las diferencias de decisión entre ambas aproximaciones, o lo que es lo mismo calcular el grado de concordancia respecto a la aceptación/rechazo de la Ho entre ellas. A tal fin, hemos calculado la Kappa de Cohen, un estadístico de concordancia que corrige el azar, y que siempre que sea posible se prefiere a los porcentajes simples de acuerdo (Bakeman y Gottman, 1989:109). Podríamos haber utilizado un contraste de hipótesis, pero predicando con el ejemplo hemos preferido una medida que nos cuantifique la magnitud de la asociación. La fórmula utilizada ha sido la clásica propuesta por Cohen (1960):

$$\kappa = \frac{Po - Pe}{1 - Pe}$$

donde **Po** es la proporción de concordancia observada realmente - el acuerdo simple entre observadores - y **Pe** la concordancia observada por azar (sumatorio del producto de frecuencias marginales):

$$Po = \frac{\text{nº acuerdos}}{\text{nº total observaciones}}$$

$$Pe = \sum pi * pj$$

Los cálculos manuales e informáticos de la magnitud de la asociación son los siguientes:

		SE		
	Ho	**RECHAZAR**	**ACEPTAR**	**pj**
SS	**RECHAZAR**	concordancia **4**	Error tipo II **11**	15
	ACEPTAR	Error tipo I **0**	concordancia **5**	5
	pi	4	15	20

Tabla 3. Matriz de confusión entre la aproximación de la SE. y SS. respecto a la decisión de la Ho

Po = 4 + 5 /20 = .45
Pc = (4 / 20 * 15 / 20) + (16 / 20 * 5 / 20) = .35
k = .45 – .35 / 1 – 0.35 = **.15**

CÁLCULO INFORMÁTICO

		SE		**Total**
SS	**Ho**	**RECHAZAR**	**ACEPTAR**	
	RECHAZAR	concordancia **4**	Error tipo II **11**	15
	ACEPTAR	Error tipo I **0**	concordancia **5**	5
Total		4	16	20

Tabla 4. Tabla de contingencia SE * SS

		Valor	**Error típ. asint.**	**T aproximada**	**Sig. aproximada**
Medida de acuerdo	**Kappa**	**0,154**	0,204	1,291	0,197
N de casos válidos		20			

Tabla 5. Valores de la magnitud de la asociación entre SE y SS

Como podemos apreciar, la magnitud de la asociación entre ambas aproximaciones es muy baja, tanto en el cálculo manual como informático, donde el resultado obtenido es idéntico (k = 0.15). Existen numerosos estándares para interpretar los valores de la Kappa de Cohen (Landis y Koch, 1977; Fleiss, 1981...), sin embargo, en lugar de una etiqueta interpretativa, Fleiss, Cohen y Everitt (1969) propusieron estimar el error típico asintótico mediante la ecuación de Fleiss (1981) y Hanley (1987), citados por Ato y López (1996):

$$\sigma\kappa = \sqrt{\frac{Pc + Pc^2 \quad - \sum pi * pj * (pi + pj)}{(1 - Pc) * \sqrt{N}}}$$

y probar la hipótesis nula Ho: k = 0 con:

$$Z = \kappa \; / \; \sigma_\kappa \cong N(0,1)$$

En nuestro estudio los datos manuales han sido los siguientes:

$$\sigma_\kappa = \sqrt{\frac{0{,}35 + 0{,}35^2 - 0{,}3525}{(1 - 0{,}35) * \sqrt{20}}} = \mathbf{0{,}203}$$

$$Z = 0{,}15 \; / \; 0{,}203 = \mathbf{0{,}738}$$

Tanto estos resultados como los que ofrece el paquete estadístico SPSS v. 10.0 son idénticos y revelan la aceptación de la Ho, es decir, denotan que el grado de acuerdo es mínimo o también que la magnitud de la asociación es inapreciable:

a) A través del cálculo manual la Z/empírica = 0,738, un valor que se asienta en la región de aceptación de Ho con cualquiera de las habituales α´s utilizadas:

Lateralidad / **Nivel α**	**Una cola (unilateral)**	**Dos colas (bilateral)**
0.05	**Z =1.64**	**Z=1.96**
0.01	**Z=2.33**	**Z=2.54**

Tabla 6. Zetas asociadas a los niveles alfas habituales dependiendo de la lateralidad asumida

b) Mediante el análisis paquete informático la significación estadística aproximada es p=0.197, y revela igualmente la aceptación de Ho.

3. CONCLUSIONES

En definitiva, hemos valorado mediante un ejemplo práctico extraído de la realidad educativa el poco acuerdo que puede llegar a existir entre el uso de la SE y la SS a la hora de decidir si las diferencias son o no son significativas. Sorprende, no obstante, que los resultados obtenidos en este estudio hayan sido opuestos a la tendencia existente basada en rechazar la Ho cuando ésta es verdadera: afirmar la existencia de diferencias cuando en realidad no las hay (error tipo I o α). Por el contrario, y si situamos como criterio de verdad/falsedad la SS, la tendencia en este estudio ha sido otra, afirmar que no hay diferencias en la percepción de los problemas relacionados en mayor o menor medida con la implantación de los nuevos planes de estudio entre los sujetos consultados comparados por binomios, cuando en realidad sí las hay (error tipo II o ß). Así, vemos que el procedimiento de la SE puede conducir no sólo a la comisión del error tipo I, sino también, y como es nuestro caso con N´s bajos, a la aparición del error tipo II. Afortunadamente, nuestro campo de investigación es el pedagógico. Imagínese el lector cuán funestas consecuencias acarrearía la comisión de este tipo de error (error tipo II) en campos como el biomédico o sanitario: no es lo mismo afirmar que dos agentes perciben de manera convergente un problema educativo, cuando es al contrario, que diagnosticar como saludable a alguien que sufre cierto grado de morbilidad. Que cada uno saque sus propias conclusiones sobre el asunto, y que además tenga en cuenta que la investigación educativa y social, en general, está llena de situaciones en que la SE debería haber sido sustituida o por lo menos completada/sopesada por la SS.

REFERENCIAS BIBLIOGRÁFICAS

ASHER, W. (1983). The role of statistics in research. *Journal of Experimental Education*, 61 (4), pp. 388-393.

ATO GARCÍA, M. y LÓPEZ GARCÍA, J.J. *(*1996). *Análisis estadístico para datos categóricos*. Madrid: Síntesis.

BAKAN, D. (1967). The test of significance in psychological research. *Psychological Bulletin*, 66, pp. 423-437.

BAKEMAN, R. & GOTTMAN, J.M. *(*1989). *Observación de la interacción. Introducción al análisis secuencial*. Madrid: Morata.

BARTOLOMÉ, M. (1984). La pedagogía experimental, en SANVIVENS, A (dir): *Introducción a la Pedagogía*, pp. 381-404. Barcelona: Barcanova.

BOTELLA, J y BARRIOPEDRO, M. I. (1995). Análisis de datos, en FERNÁNDEZ-BALLESTEROS, R.: *Evaluación de Programas*. pp. 173-207. Madrid: Síntesis.

CHOW, S.L. (1998). The significance test or effect size? *Phichological Bulletin*, 103, pp. 105-110.

COHEN, J. A. (1960). coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 pp. 37-46.

CRONBACH, L.J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 34, pp. 513-531.

DE LA ORDEN, A. Y MAFOKOZI, J. (1997). Implicaciones de algunos planteamientos epistemológicos post-positivistas en la investigación educativa. *Bordón*, 49 (4), pp. 347-358.

DENDALUZE, I. (1998). Algunos retos metodológicos. *Revista de Investigación Educativa*. V 16, 1 pp. 7-27.

FERNÁNDEZ CANO, A. (1995). *Métodos para evaluar la investigación en psicopedagogía*. Madrid: Síntesis.

FISHER, R.A. (1959). *Statistical methods and scientific research*. NY. Hafner

FLEISS, J.L. (1989). *Statistical methods for rates and proportions*. NY. Wiley.

FLEISS, J.L.; COHEN, J. y EVERITT, B.S. (1969). Large-sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 72, pp. 295-300.

GUTTMAN, L. The illogic of statiscial inference for cumulative science. *Applied Stochastic Models and Data Analysis*, 1 pp. 3-10. 1985.

HARCUM, E.R. (1989). The highly inappropiate calibrations of statistical significance. *American Psychologist*, 4, pp. 49-64.

HUBERTY, C.J. (1987). On statistical significance. *Educational Researcher*, 16 (8), pp. 4-9.

LANDIS, J.R. y KOCH, G.G. (1977). An aplication of hierchical kappa-type statistics in the assement of majority agreement among multiple observer. *Biometrics*, 33, pp. 363-374.

LEÓN, O. (1984). El uso del término significativo en los informes experimentales. *Revista de Psicología General y Aplicada*, 39, pp. 455-469.

MEEHL, P.E. (1967). Theory-testing in psychology and physics: *A methodological paradox. Philosophy of science*, 34, pp. 103-115.

MORRISON, D.E. y HENKEL, R.E. (1970) (edit.) *The significance test controversy: A reader. Hawthorne*, NY. De Gruyter Aldine.

NEYMAN, J. Y PEARSON, E. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Part I and II. *Biometrika*, 20A, pp. 175-240/262-194

OAKES, M. (1986). Statistical Inference: *A commentary for the social and behavioral sciences*. NY. Wiley.

PICK, S. y LÓPEZ, A.L. (1994). *Cómo investigar en ciencias sociales*. México. Trillas. (5ª edición).

ROSNOW, R.L. y ROSENTHAL, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 30, pp. 1165-1168.

SCRIVEN, M. (1988). Philosophical inquiry methods in Education (Section III), en JAEGER, R.M. (edit): *Complementary methods for research in education*. Washington, D.C. AERA, pp. 131-148,

SEOANE, J. y otros. *(*1987*). Psicología Matemática I.* Madrid. UNED.

SHULMAN, L.S. Paradigms and research programs in the study of teaching, en WITROCK, M. C. (ed): *Handbook of research on teaching*. N.Y. Mc Millan.

SIEGEL, S. *(*1991). *Estadística no paramétrica aplicada a las ciencias de la conducta.* México. Trillas.

SIGNORELLI, A. (1974). Statistics: Tool or master of the psychologist?. *American Psychlogist*, 29, pp. 774-777.

SKINNER, B.F. A. (1956). case history in scientific method. *American Psychologist*, 11, pp. 221-223.

THOMPSON, B. (1989). Statistical significance, result importance, and result generalizabilty: Three note worthy but somewhat different issues. *Measurement and Evaluation in Couseling and Development*, 22, pp. 2-5.

WALKERS, J. C. y EVERS, C. W. (1990). En THOMAS, R.M. *The encyclopedia of human development*, Oxford. Pergamon Press.