Análisis del Proceso de Construcción de un Cuestionario sobre Probabilidad Condicional. Reflexiones desde el Marco de la TFS

Carmen Batanero y Carmen Díaz, Universidad de Granada
batanero@ugr.es, mcdiaz@ugr.es

RESUMEN

Describimos brevemente el proceso de elaboración de un cuestionario de evaluación, y lo analizamos desde el punto de vista de la TFS. La finalidad es reflexionar sobre las instituciones y procesos de muestreo implicado, así como sobre las posibilidades de generalización y criterios de idoneidad de las tareas e instrumentos de evaluación.

1. INTRODUCCIÓN

La construcción de instrumentos de evaluación es habitual en la investigación en educación, donde se siguen en mayor o menor medida las normas metodológicas habituales en psicometría. Desde el punto de vista del marco teórico de las TFS, un instrumento de evaluación tiene como finalidad proporcionar información sobre los *significados personales* de un grupo de estudiantes sobre un objeto o un grupo de objetos matemáticos dados y la investigación encaminada a la construcción de estos instrumentos o el análisis de las respuestas a los mismos entraría en el campo de la *semiometría* (Godino, 1999; 2002). La semiometría contempla lo que podemos describir como *estática de significados sistémicos*, esto es, la caracterización de la trama de las funciones semióticas (o al menos una muestra representativa de tal trama) en las cuales un objeto se pone en juego en un contexto y circunstancias fijadas. La "medida" de tales significados (sistemas de prácticas) tendrá un carácter cualitativo y será relativa a una persona, institución, contexto fenomenológico y momento temporal especificado). Aunque el objeto de estudio de la evaluación es el significado personal, este proceso no puede dejar de lado la faceta institucional de los objetos matemáticos, que sirve de pauta de comparación con los significados personales evaluados.

Por otro lado, ni el significado holístico o global de un objeto de enseñanza ni siquiera los significados pretendidos o implementados en una enseñanza sobre dicho objeto pueden ser abarcados en un solo instrumento de evaluación. Tampoco el significado personal del alumno puede ser explicitado completamente en las respuestas a tareas o la observación de su actividad durante una prueba. En este trabajo describiremos brevemente el proceso seguido en la construcción de un cuestionario dirigido a evaluar la comprensión de la probabilidad condicional por estudiantes de Psicología. La finalidad principal es reflexionar sobre la complejidad de la función evaluadora, los diferentes niveles y tipos de significados personales e institucionales involucrados y sobre la información que diferentes tipos de análisis psicométricos pueden proporcionar respecto a las posibilidades de generalización en la investigación educativa.

1.1. EL CONTEXTO DE LA INVESTIGACIÓN

La construcción del cuestionario citado forma parte de una investigación realizada en el Departamento de Psicología Social y Metodología de las Ciencias del Comportamiento de la Universidad de Granada (Díaz, 2004). El concepto de probabilidad condicional es fundamental en las aplicaciones de la Estadística, porque permite incorporar cambios en nuestro grado de creencia sobre los sucesos aleatorios a medida que adquirimos nueva información. Es también un concepto teórico básico requerido en el estudio de la inferencia estadística, tanto clásica como bayesiana, así como en el estudio de la asociación entre variables, la regresión y los modelos lineales.

En el terreno profesional e incluso en la vida cotidiana, la toma de decisiones acertadas en situaciones de incertidumbre se basa en gran medida en el razonamiento condicional.

1.2. CONSTRUCTOS Y VARIABLES

Al tratar de evaluar la comprensión sobre un cierto concepto de un grupo de alumnos, hemos de tener en cuenta que es un *constructo inobservable* (León y Montero, 2002), por lo que sus características deben ser inferidas de las respuestas de los alumnos. Un constructo es un atributo psicológico que caracteriza los comportamientos de los individuos y nos permite explicar patrones de comportamiento. Sólo pueden ser observados indirectamente, están sujetos al cambio y sólo los comprendemos vagamente, de aquí la dificultad de su evaluación que llevamos a cabo mediante alguna *variable* observable, por ejemplo, la puntuación en un cuestionario (Osterlind, 1989). Generalmente hay más de una posible forma de definir el constructo, cuya definición se realiza a dos niveles:

- *Definición semántica*: en términos de comportamientos observables o reglas de correspondencia entre el constructo y la conducta.
- *Definición sintáctica:* en términos de las relaciones lógicas o matemáticas del constructo con otros constructos o variables dentro de un marco teórico.

En este trabajo nos limitaremos a la definición semántica, que considera la especificación detallada de la variable de interés (en este caso la comprensión de la probabilidad condicional). Desde nuestro marco teórico, las especificaciones del contenido podrían describirse como tipos de prácticas operatorias y discursivas (Godino, 2003) asociadas al objeto "probabilidad condicional". La diferenciación entre constructos y variables también se recoge en las TFS, puesto que se diferencia entre el dominio de las ideas u objetos abstractos (personales e institucionales) y el dominio de los significados o sistemas de prácticas de donde emergen tales objetos inobservables (Godino, 1999), lo que permite plantear con nitidez la dificultad del problema de la evaluación.

2. SIGNIFICADO DE REFERENCIA DE LA PROBABILIDAD CONDICIONAL

Nuestro primer paso ha sido elaborar un procedimiento para identificar, mediante una definición semántica precisa, el constructo (Martínez Arias, 1995). Para dotar de una mayor objetividad a esta definición de nuestra variable, nos hemos basado en un análisis de contenido de una muestra de libros de texto que utilizan los alumnos de Psicología en las asignaturas de análisis de datos. Además hemos tenido en cuenta los errores y dificultades señalados en las investigaciones previas sobre la probabilidad condicional. Con todo ello delimitamos el *significado de referencia* en nuestro estudio. Hacemos notar que este es un significado parcial, puesto que el objeto probabilidad condicional tiene un significado más completo, si se tienen en cuenta los elementos aportados a dicho significado desde la matemática (por ejemplo, en nuestro estudio no trataremos el concepto de intercambiabilidad), la historia (sucesivas concepciones históricas de la probabilidad condicional y los campos de problemas que las originaron), psicología y didáctica. Todo ello constituiría el *significado holístico o global* del concepto.

2.1. INVESTIGACIONES PREVIAS

En primer lugar partimos de las principales investigaciones relacionadas con la comprensión de las ideas de probabilidad condicional e independencia, tanto en el campo de la Psicología, como en el de la Educación, recopilando, además, los ítems usados en las mismas, que serían la base posterior de la construcción de un banco de ítems. Podemos clasificar las investigaciones encontradas en los apartados siguientes:

- *Comprensión intuitiva de la probabilidad condicional y sus relaciones con la probabilidad simple y dependencia* (Maury, 1985, 1986; Kelly y Zwiers, 1986; Totohasina, 1992; Sánchez, 1996).
- *Condicionamiento y causación:* La existencia de una relación condicional indica que una relación causal es posible, pero no segura. Desde el punto de vista psicológico, la persona que evalúa una probabilidad condicional percibe en forma diferente las relaciones causales y diagnósticas (Tversky y Kahneman, 1982a). La relación de causalidad también se asocia, a menudo, con la secuencia temporal (Falk, 1986; Gras y Totohasina, 1995).
- *Intercambio de sucesos en la probabilidad condicional* (Eddy, 1982; Falk, 1986; Batanero y cols., 1996).
- *Confusión de probabilidad condicional y conjunta* (Pollatsek, Well, Konold y Hardiman, 1987; Einhorn y Hogarth, 1986; Ojeda, 1995; Tversky y Kahneman, 1982b).
- *Situaciones sincrónicas y diacrónicas:* Ojeda (1995).*Razonamiento bayesiano* (Kahneman & Tversky, 1982c; Bar-Hillel, 1983; Totohasina, 1992; Teigen, Brun & Frydenlund, 1999)
- *Influencia del formato y los datos* (Pollatesk y cols.,1987; Fiedler, 1988; Gigerenzer, 1994)
- *Enseñanza de la probabilidad condicional (*Sdlemeier, 1999; Martignon & Wassner, 2002).

Esta revisión nos permitió comprobar que las investigaciones se habían centrado en puntos aislados de la comprensión del concepto y sugirió la necesidad de construir un cuestionario comprensivo. Por otro lado, se enriquece el significado puramente matemático del concepto, al tener en cuenta aspectos psicológicos involucrados, tales como la falacia de la conjunción (Tversky y Kahneman, 1982b) o la falacia del eje temporal (Falk, 1986).

2.2. ANÁLISIS DE CONTENIDO DE LIBROS DE TEXTO DE ESTADÍSTICA PARA PSICÓLOGOS

El estudio de los libros de texto es una forma –limitada- de acercarnos al significado institucional de la probabilidad condicional en la institución "análisis de datos en psicología", es decir en los cursos universitarios de análisis de datos para este tipo de estudiantes. El análisis de contenido se basa en la idea de que las unidades del texto pueden clasificarse en un número reducido de categorías (Weber, 1985). Sirve para efectuar inferencias mediante la identificación sistemática y objetiva de las características específicas de un texto (Ghiglione y Matalón, 1989). El procedimiento seguido consistió en elaborar un listado con las 31 universidades españolas en las que se imparte la licenciatura de Psicología y solicitar a los directores de los correspondientes departamentos el programa y bibliografía recomendada. Conseguimos repuestas de 23 de estas universidades. De un total de 79 libros diferentes recomendados de análisis de datos 20 eran citados por 4 o más universidades. Trece de ellos incluían el tema de probabilidad condicional y fueron analizados. Además se incluyeron otros 5 libros de orientación bayesiana.

2.3. ESPECIFICACIÓN DE LA VARIABLE

El análisis realizado de los libros y las investigaciones previas sirvió para elaborar la tabla de especificaciones de nuestro cuestionario, que se presenta en la Tabla 1. Aunque al elaborar esta tabla sólo se diferenció el nivel conceptual como procedimental, así como los cuatro niveles superiores en la taxonomía de Bloom,

podríamos haber diferenciado los diferentes elementos de significado considerados por Godino (2003).

Tabla 1. Especificaciones del contenido del cuestionario

	Contenido	Comprensión	Aplicación	Análisis	Síntesis
Conocimiento conceptual	1. Definición de la probabilidad condicional	x	x		
	2. Reconocer que la probabilidad de P(A) >0 para poder definir P(B/A)	x			
	3. Reconocer que una probabilidad condicional cumple los axiomas.	x			
	4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral	x	x		
	5. Distinguir probabilidad condicional con inversa	x	x		
	6. Distinguir probabilidad conjunta, condicional y simple	x	x		
	7. Probabilidad conjunta menor que probabilidad simple	x	x		
	8. Distinguir sucesos independientes, dependientes y mutuamente excluyentes	x	x		
Conocimiento procedimental	9. Calcular una probabilidad condicional dentro de un único experimento			x	x
	10. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo con reposición			x	x
	11. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo sin reposición			x	x
	12. Resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples			x	x
	13. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo			x	x
	14. Distinguir situación condicional, causal y diagnóstica			x	x
	15. Resolver correctamente problemas en situaciones diacrónicas			x	x
	16. Resolver correctamente problemas en situaciones sincrónicas			x	x
	17. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes			x	x
	18. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes			x	x
	19. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total)			x	x
	20. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes)			x	x

En nuestro caso, estamos considerando los conceptos y propiedades (conocimiento conceptual) así como los problemas y algoritmos (conocimiento procedimental), aunque no se diferencian entre si este tipo de elementos. No hemos prestado atención especial al lenguaje o a los argumentos, aunque por supuesto quedan implícitamente evaluados en los ítems de respuesta abierta, que se podrían reanalizar para estudiar la comprensión de este tipo de elementos.

3. PROCESO DE SELECCIÓN DE ÍTEMS

Un ítem de un cuestionario es una unidad de medida que consta de un estímulo y una forma prescriptiva de respuesta y su fin es inferir la capacidad del examinado en un cierto constructo (habilidad, rasgo, etc.), proporcionando datos cuantificables sobre la persona que lo completa (Osterlind, 1989). Desde el punto de vista de la TFS, y aún teniendo en cuenta la naturaleza esencialmente compleja del significado de los objetos matemáticos, en el análisis de las actuaciones de los alumnos nos interesa con frecuencia fijar la atención en procesos interpretativos específicos y en las dificultades inherentes a los mismos, por lo que en la respuesta a cada ítem podríamos tener en cuenta *significados parciales* o incluso *elementales* de la probabilidad condicional (Godino y Batanero, 2003).

Al considerar el número total de ítems, hemos tratado de cubrir adecuadamente el contenido y asegurar una fiabilidad satisfactoria (Millman y Greene, 1989), teniendo en cuenta la restricción de la posible longitud total del test. Usamos tanto ítems de opciones múltiples como de respuesta abierta y comenzamos con un conjunto inicial de unos 50 ítems (2-3 por cada especificación de contenido).

Una vez concluida la planificación del cuestionario, realizamos una selección de los ítems que constituirían el cuestionario piloto, siguiendo los dos sistemas sugeridos en Osterlind (1989):

- El análisis a partir de un juicio requiere pedir a una serie de expertos que valoren los ítems particulares, de acuerdo con algunos criterios.
- La valoración numérica requiere que los ítems se administren a una muestra de sujetos y se basa en el estudio de una serie de indicadores estadísticos de los mismos.

3.1. JUICIO DE EXPERTOS

Millman y Greene (1989) indican que el "experto" lo define el propósito del instrumento y que el grupo elegido de expertos ha de representar una diversidad relevante de capacidades y puntos de vista. En nuestro caso, fueron seleccionados en base a su conocimiento experto de probabilidades y más particularmente de probabilidad condicional, así como a la experiencia de investigación sobre el tema. Participaron 9 investigadores en didáctica de la estadística, tanto españoles como iberoamericanos. Nuestro objetivo era doble:

- Establecer un consenso sobre la tabla de especificaciones del instrumento, decidiendo cuales especificaciones del contenido eran relevantes para los propósitos del instrumento. De este modo se reforzarían los resultados obtenidos del análisis de contenido de los libros de texto.
- Establecer un consenso de opiniones de los expertos sobre cómo cada ítem particular se ajusta bien para evaluar el contenido específico para el cuál ha sido diseñado que sirviesen como base para elegir los ítems definitivos.

Se proporcionó a cada uno de estos expertos un cuestionario en que se les pedía, para cada unidad de contenido y para cada uno de los ítems asociados a la misma su grado de acuerdo (en una escala 1 a 5) sobre su adecuación a los fines de nuestra evaluación. Un ejemplo de la estructura y contenido del cuestionario a expertos se muestra en la Figura 1.

Figura 1. Ejemplo de contenido y estructura del cuestionario a expertos

Contenido 1: Definición de la probabilidad condicional.

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 2. ¿Qué quiere decir la expresión "la probabilidad condicional de A dado B es 1/4"?

a) En la cuarta parte de los experimentos obtenemos A y B simultáneamente
b) A ocurre la cuarta parte de las veces en que ocurre B
c) B ocurre la cuarta parte de las veces en que ocurre A
d) A o B ocurren la cuarta parte de las veces

	1: Nada	2	3	4	5: Mucho
El contenido "Definición de la probabilidad condicional" es relevante					
El ítem 1 es adecuado para este contenido					
El ítem 2 es adecuado para este contenido					

3.2. PRUEBAS PREPILOTOS DE ITEMS

Los ítems fueron divididos en cuatro cuestionarios, con objeto de que el número total de ítems a completar en una sesión por un mismo grupo de alumnos no fuera excesivamente largo, de modo que tuviesen tiempo suficiente para responder. Participaron en las pruebas pre – piloto dos grupos diferentes de alumnos, de primer año en la Licenciatura de Psicología (en total 157 alumnos, con una nota media de acceso en selectividad de 7,49). Los porcentajes de estudiantes con diferentes tipos de bachillerato fueron los siguientes: ciencias 27'6%, letras 72,4%. Se trata de sujetos voluntarios, como es frecuente en investigaciones en ciencias sociales.

Tabla 2. Resultados en pruebas pre-piloto

Ítem	n	porcentaje correcto	Observaciones
1	49	75.5	
2	52	71.2	No eligen el distractor d
3	49	24.5	
4	52	73.1	
5	157	24.8	Domina el distractor c
6	49	36.7	
7	52	15.4	Domina el distractor d
8	49	36.7	
9	52	a) 65.4; b) 46.2	
10	49	a) 46.9; b) 38.8	
11	157	47.1	Domina el distractor c
12	52	a) 76.9;b) 44.2; c) 65.4	
13	49	a) 24.5; b) 71.4; c) 46.9	
14	81	37.0	Domina el distractor c
15	76	27.6	Domina el distractor c

3.3. SELECCIÓN DE ÍTEMS PARA EL CUESTIONARIO

Una vez finalizadas las fases anteriores, seleccionamos los ítems que compondrían el cuestionario piloto, con el siguiente proceso:

- Primeramente se desecharon aquellas especificaciones de contenido en las que el grado de acuerdo sobre su relevancia no fuese suficientemente elevado y consensuado.
- Para aquellos contenidos con alto grado de acuerdo respecto a su relevancia se examinaron los ítems. Dichos ítems no tenían *idoneidad epistémica* para el elemento particular de significado evaluado. Se desecharon los ítems que no fuesen altamente valorados por los expertos. De entre los ítems bien valorados se eligió el que hubiese presentado menor índice de dificultad en las pruebas pre-piloto de ítems. Los ítems excesivamente difíciles no tenían *idoneidad cognitiva.* Algunos ítems también se desecharon porque algún distractor fue elegido mayoritariamente (mientras que lo ideal es que las respuestas se distribuyan homogéneamente entre distractores) o porque el contexto muy marcado –por ejemplo campeonatos de tenis- distraía la atención del estudiante y forzaba la respuesta. En este caso, incluso cuando el ítem

tuviese idoneidad epistémica y cognitiva no tenía *idoneidad didáctica,* más específicamente no es adecuado a la tarea evaluadora del conocimiento pretendido.

4. PRUEBAS DEL CUESTIONARIO PILOTO

4.1. SIGNIFICADO EVALUADO EN EL CUESTIONARIO

Una vez elegidos los ítems que formarían el cuestionario (18) se analizó su contenido primario y secundario. Por ejemplo, el ítem siguiente (número 4 en el cuestionario) fue diseñado para evaluar el contenido primario *"distinguir sucesos dependientes, independientes y mutuamente excluyentes"*. Pero al analizarlo aparecen otros contenidos secundarios, ya que el alumno debe también "*resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en experimentos independientes"* para reconocer que el distractor d) es incorrecto.

Ítem 3. Se extrae una carta al azar de una baraja americana: sea A el suceso "se extrae un trébol" y B el suceso "se extrae una reina" ¿Los sucesos A y B son independientes?

a) Sí, en todos los casos.
b) No son independientes porque en la baraja hay una reina de tréboles.
c) Sólo si sacamos primero una carta para ver si es reina y se vuelve a colocar en la baraja y luego sacamos una segunda carta y para mirar si es trébol.
d) No, porque P(reina de trébol)= P(reina) x P(trébol)

Los otros distractores tratan de detectar diferentes errores descritos en la literatura de investigación sobre probabilidad condicional:

- En el distractor b) se trata de detectar la posible confusión entre sucesos excluyentes y sucesos independientes.
- En el distractor c) se trata de detectar la posible creencia errónea que sólo pueden ser independientes los sucesos de experimentos diacrónicos.
- En el distractor d) se presenta la definición correcta de la regla del producto, junto con la afirmación incorrecta que esta regla no se cumple en el caso de sucesos independientes. Precisamente en este caso se cumple porque los sucesos son independientes y tratamos de ver si el alumno detecta el error en la afirmación presentada en el distractor.

Este análisis para cada uno de los ítems permitió comprobar que todas las unidades de contenido especificadas en la tabla 1 quedaban cubiertas al menos dos veces en el cuestionario. En esta tabla representamos, por tanto el significado de la probabilidad condicional *evaluado* por el cuestionario, que es más restringido que el significado holístico y en ciertos aspectos también que el socio-profesional en la institución de enseñanza – evaluado en el análisis de textos. Pero incluye elementos psicológicos, tomados de las investigaciones previas, que no se tienen en cuenta en la enseñanza y por tanto son nuevos respecto al significado institucional socio-profesional.

4.2. ANÁLISIS DE ÍTEMS

Finalizado el instrumento piloto, decidimos realizar pruebas del mismo, con objeto de obtener información empírica sobre las características de esta primera versión del instrumento y comprobar que era útil para los objetivos pretendidos. Al mismo tiempo deseábamos analizar sus limitaciones, con objeto de identificar aquellos puntos en que sería necesario continuar el trabajo para mejorarlo. Las pruebas piloto se llevaron a cabo con dos muestras de estudiantes universitarios. La primera de ella estuvo formada por 37 alumnos que cursaban 5º año de la Licenciatura de Matemáticas en la especialidad de Metodología. Se eligió a estos alumnos, debido a su alta preparación y porque el

disponer de este grupo de alumnos nos permitiría comparar si los errores más frecuentes en alumnos de psicología se repetían en alumnos con alta preparación matemática.

Análisis de respuestas

Analizamos, en primer lugar, las respuestas obtenidas en cada uno de los ítems. En los ítems de opciones múltiples (como el ítem 2) que se presenta a continuación simplemente estudiamos las frecuencias y proporciones de respuestas (Tabla 3).

Ítem 4. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

Tabla 3. Resultados en el Item 4

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) 80 %	7	18,9	10	17,5
b) 15%	6	16,2	2	3,5
c) (15/100) x (80/100)	18	48,6	17	29,8
d) 41 %	4	10,8	16	28,1
Blanco	2	5,4	12	21,1
Total	37	100,0	57	100,0

Ítem 5. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cuál es la probabilidad de que ninguno de los dos números sea un 6?

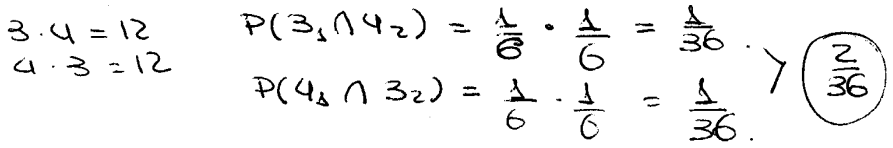
En los ítems de respuesta abierta se ha puntuado de acuerdo a la mayor o menor completitud. Por ejemplo, en el ítem 3 se ha seguido el siguiente criterio: 0:No se responde, o se responde incorrectamente; 1: Identifica los casos favorables o posibles, pero no resuelve el problema; 2) Identifica los casos favorables y posibles, plantea el problema pero comete algún error y 3) Resuelve el problema correctamente, como en el caso siguiente: *"{(2,6), (3,4), (6,2), (4,3)}; 2/4=1/2= 0,5"*.

Tabla.4. Resultados en el ítem 5

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	1	2,7	2,7	7	12,3	12,3
0	13	35,1	37,8	23	40,4	52,7
1	4	10,8	48,6	12	21,1	73,8
2	7	18,9	67,5	9	15,8	89,6
3	12	32,4	100,0	6	10,5	100,0
Total	37	100,0		57	100,0	

Aunque no los hemos analizado explícitamente, estos ítems de respuesta abierta permiten identificar los elementos de significado usados por el alumno, así como sus conflictos semióticos. Un ejemplo es el siguiente en que el alumno ha identificado los casos favorables, identifica la independencia de los sucesos y la importancia del orden pero no identifica el problema como de probabilidad condicional, sino lo confunde con otro de probabilidad simple. Por ello, encuentra la probabilidad de cada caso, aplicando

la regla del producto (correctamente) y suma las probabilidades (aplicando el axioma de la unión). El alumno muestra un razonamiento probabilístico bueno, y ha aplicado un gran número de elementos de significado, pero aparece un *conflicto semiótico* al no interpretar correctamente el enunciado del problema.



.

Seguidamente se analizaron las características psicométricas de los ítems, dificultad y discriminación. Puesto que el índice de dificultad es una proporción (la proporción de alumnos que superan el ítem) calculamos adicionalmente los intervalos de confianza de dicha proporción con la fórmula habitual, mediante la hoja de cálculo Excel. Para el cálculo de la proporción de aciertos en los ítems de respuestas abiertas en los que, además de la respuesta totalmente correcta (máxima puntuación), también se han considerado las respuestas en que el razonamiento e identificación de los datos sea correcta, aunque haya algún error de cálculo.

Una aportación metodológica es el uso de inferencia bayesiana para mejorar la estimación de los índices de dificultad, utilizando el programa Le Bayesien (Lecoutre, 1996) que realiza el cálculo bayesiano con proporciones (Bernard, 1998). Esta estimación se ha realizado con dos tipos de distribución a priori:

- Distribución a priori no informativa, es decir, sin tener en cuenta la información previa (Lecoutre, 1996).
- Utilizando resultados del ítem en las pruebas pre-piloto, siguiendo uno de los métodos recomendados en Serrano (2003), que consiste en tener en cuenta los éxitos y fracasos en la prueba anterior para definir la distribución a priori.

En ambos casos calculamos los intervalos de credibilidad y los comparamos con los intervalos de confianza obtenidos en la aproximación clásica. Cuando se dispone de información previa, los intervalos obtenido son más precisos y además nos proporciona una probabilidad epistémica (es decir de que el parámetro esté en este intervalo), lo que no se logra en inferencia clásica.

Tabla 4. Estimación de índices de dificultad

Ítem	Índice observado en prueba piloto	Intervalo de confianza (95%)	Índice corregido	(Éxitos/fallos) en Prueba pre-piloto	Intervalo de credibilidad bayesiano (95%) según d. a priori	
	(n=57)				No informativa	Informativa
I1	0,860	.7667-.9526	0,853	(37, 12)	.753-.931	0,729-0,877
I2	0,720	.5990-.8396	0,716	No probado	.594-.823	---
I3	0,281	.1604-.4010	0,284	(23, 134)	.177-.406	.135-.238
I4	0,175	.0736-.2773	0,171	(26, 131)	.094-0.289	.123-.223
I5	0,825	.7227-.9264	0,819	(45, 31)	.711-.906	.610-.765
I6A	0,877	.7893-.9651	0,871	(23, 26)	.774-.943	.596-.771
I6B	0,298	.1758-.4207	0,299	(19, 30)	.192-.425	.261-.433
I6C	0,421	.2889-.5532	0,422	(12, 37)	.264-.550	.255-.433
I6D	0,421	.2889-.5532	0,422	(35, 14)	.264-.550	.462-.649

Discriminación

También hemos calculado los índices de discriminación para cada uno de los ítems por dos sistemas diferente:

1. Como correlación entre el ítem y la puntuación total del cuestionario. La mayor parte de los ítems tienen una correlación positiva; pero a veces no es

suficientemente significativa lo que da idea de un constructo sistémico más que unidimensional.

2. Mediante el estudio de la diferencia en proporción de aciertos al ítem en los estudiantes de psicología respecto a un grupo de mayor preparación (estudiantes de matemáticas, con alta formación en matemáticas y estadística), aplicando el test usual de diferencia de proporciones. Resaltamos que en algunos ítems los estudiantes de matemáticas no obtienen mejores resultados que sus compañeros, posiblemente porque los sesgos de razonamiento no se corrigen con la instrucción exclusivamente formal.

Tabla 5. Dificultad y discriminación

	Índice Dificultad		Índice Discriminación n = 94		
	Matemáticas	Psicología	Diferencia grupos		Correlación corregida con total
	n = 37	n =57	Valor diferencia	Valor tipificado	
I1	0,784	0,860	-0,076	-0,93	***,687
I2	0,784	0,720	0,064	0,71	**,297
I3	0,108	0,281	-0,173	-2,21	-,202
I4	0,081	0,175	-0,094	-1,39	-,036
I5	0,811	0,825	-0,014	-0,17	,107
I6A	0,946	0,877	0,069	1,21	*,205
I6B	0,459	0,298	0,161	1,58	*,215
I6C	0,730	0,421	***0,309	3,15	**,371
I6D	0,730	0,421	***0,309	3,15	**,345

* p=0.05; ** p= 0.01; *** p=0.001

4.3. FIABILIDAD, GENERALIZABILIDAD Y VALIDEZ

Una vez realizado el análisis de ítems, hemos tratado de buscar algunas evidencias de la fiabilidad y validez de la prueba, siendo conscientes que, al estar todavía en las fases iniciales de la construcción, estas evidencias son sólo provisionales y deberán ser completadas en el futuro. Como sugiere Godino (1999), la solución de los problemas de investigación, con criterios de calidad científica, precisa realizar un trabajo sistemático y disciplinado que garantice la validez y fiabilidad de las afirmaciones pretendidas, esto es, debe estar guiada por una metodología adecuada de investigación y por instrumentos teóricos adaptados a las peculiaridades de la investigación requerida.

Aproximación a la fiabilidad

Puesto que estamos en un proceso de evaluación en educación, nos proponemos realizar inferencias sobre conceptos abstractos a partir de indicadores empíricos, más precisamente, relacionar los conocimientos de los alumnos sobre un concepto con sus respuestas a los ítems del cuestionario (Thorndike, 1989). La medida siempre produce un cierto error aleatorio, pero dos medidas del mismo fenómeno sobre un mismo individuo suelen ser consistentes. La fiabilidad es esta tendencia a la consistencia o precisión del instrumento en la población medida. Se define como correlación entre las puntuaciones verdadera y observada (Martínez Arias, 1995).

El *coeficiente de fiabilidad* es un indicador de la fiabilidad teórica de las puntuaciones observadas, en el sentido de proporcionar un valor numérico que indica el *grado de confianza* que podíamos tener en dichas puntuaciones como estimadores de las puntuaciones verdaderas de los sujetos. Este coeficiente de fiabilidad es un valor teórico que debe ser estimado por algún procedimiento empírico, a través de las respuestas de un grupo de sujetos a un conjunto de ítems (Carmines y Zeller, 1979). Entre los diversos procedimientos para el cálculo del estimador del coeficiente de fiabilidad hemos tomado el coeficiente Alfa de Cronbach (que se reduce al de Kuder- Richardson para ítems dicotómicos). Este coeficiente nos refleja el grado en el que covarían los ítems que

constituyen el test. Es la estimación de una *fiabilidad en el acto*. Se acerca la puntuación de una persona a la que se obtenido si tuviésemos un instrumento perfecto de medición (Martínez Arias, 1995).

El valor obtenido con el total de la muestra es de 0,7738, que corresponde a una correlación entre la puntuación observada y la puntuación verdadera de 0,879. Consideramos que el valor es razonable, dado que se trata de un cuestionario piloto que puede ser mejorado. Al calcular separadamente el coeficiente de fiabilidad para ambos grupos obtuvimos un valor de alfa = 0,6043 para el grupo de Psicología (n = 57) y un valor alfa de 0,5604 en el grupo de Matemáticas (n =37). Estos valores son moderados, como corresponde a un constructo no unidimensional, pero entra en los límites sugeridos por diversos autores; por ejemplo Santisteban (1990) indica, como limite general, 0,50.

Generalizabilidad

La teoría de la generalizabilidad (Feldt y Brennan, 1991) extiende la teoría clásica de la medición, y permite analizar diferentes fuentes de error en las puntuaciones observadas. El coeficiente de generalizabilidad se define como cociente entre la varianza verdadera en las puntuaciones de la prueba y la varianza observada que es suma de la varianza verdadera más la varianza debida al error aleatorio, mediante la expresión (1).

$$(1).....G = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}$$

Según Thorndike (1989), la varianza de error depende de cómo definimos el universo de puntuaciones verdaderas y en el análisis de generalizabilidad se consideran ciertas fuentes como parte de la varianza de error en unas condiciones y otras fuentes en otras. En nuestro caso diferenciaremos dos fuentes para el error aleatorio y calcularemos, por tanto, dos coeficientes: a) la generalizabilidad a los mismos alumnos cuando se varían los enunciados de la misma prueba, que coincide con el coeficiente alfa; por tanto para el grupo de psicólogos es igual a 0,60; b) la generalizabilidad a otros alumnos similares a los del grupo con la misma prueba fue igual a 0,90. Estos resultados indican una alta generalizabilidad de resultados usando el mismo cuestionario.

Aproximación a la validez

Reconocemos que la validación de un cuestionario es un proceso complejo de recogida y documentación de evidencia y entendemos la validez como un concepto que es unitario, ya que sus diferentes formas (validez de constructo, contenido, criterio...) deben considerarse más bien como diferentes maneras de recoger evidencias de una única noción de validez (Martínez Arias, 1995). La validez suele entenderse en relación al uso que se dé al cuestionario y la interpretación de sus puntuaciones. No validamos un test sino una inferencia o interpretación (Osterlind, 1989).

Nosotros hemos tratado de iniciar el estudio de la *validez de contenido* es decir, asegurar que el instrumento recoge una muestra representativa de los contenidos que se pretenden evaluar o que los ítems son relevantes para representar un universo temático (Carmines y Zeller, 1979). Para ello hemos realizado una planificación cuidadosa de los ítems que se incluirán en el mismo y un estudio de cómo estos ítems pueden contribuir a la medida del constructo subyacente. Para salvar la dificultad de determinar qué es un muestreo adecuado de ítems en el dominio particular, hemos descrito de antemano el dominio, sus dimensiones, facetas y objetivos (Martínez Arias, 1995).

También hemos usado una metodología adecuada para revisar la congruencia entre los ítems y las especificaciones (Osterlind, 1989). El establecimiento de la validez

mediante la determinación de la correspondencia entre ítems y objetivos se aplica especialmente a los tests de rendimiento como es el nuestro. Como es usual en este tipo de validación hemos usado el juicio de expertos, definiendo previamente el universo de observaciones admisibles, identificando expertos en el campo, pidiendo a los expertos que emparejen ítems con objetivos y resumiendo la información numéricamente.

El trabajo de construcción del cuestionario no ha finalizado. Somos conscientes de que necesitamos nuevas revisiones y pruebas con muestras más amplias que nos permitan completar el estudio de la validez, así como la mejora de la fiabilidad. Sería también preciso complementar los resultados del cuestionario con entrevistas en profundidad, que proporcionen un conocimiento más profundo de los razonamientos y conocimientos de los estudiantes.

5. REFLEXIONES DESDE EL ENFOQUE DE LA TFS

5.1. INSTITUTIONES IMPLICADAS EN EL PROCESO DE EVALUACIÓN Y SIGNIFICADOS DE LA PROBABILIDAD CONDICIONAL

El proceso de construcción y prueba de un cuestionario de evaluación descrito nos permite reflexionar sobre la complejidad de la tarea evaluadora y la diversidad de significados involucrados respecto a un mismo objeto matemático. En la descripción aparece claramente la dialéctica *institucional-personal* (Godino, 2003), puesto que la adecuación de los significados personales de los alumnos sólo puede llevarse a cabo desde una institución de referencia. Ahora bien, dicha faceta aparece en relación con diversas instituciones que intervienen en diferentes fases, que representamos esquemáticamente en la Tabla 6.

Tabla 6. Significados y procesos de muestreo en el estudio realizado

Instituciones/ sujetos involucrados	Significado de interés	Instrumento de evaluación	Proceso de muestreo	Significado evaluado
Didáctica / Matemática	Holístico			
Enseñanza (Estadística en Psicología)	Institucional (socio-profesional)	Análisis del contenido de libros de texto recomendados	23 entre 31 Universidades; 18 entre 60 libros	Referencia
Investigación en Educación Estadística	Objetivamente evaluado	Juicio de expertos Pruebas pre-piloto de ítems	9 investigadores muestras de estudiantes	Evaluado
Estudiantes de Psicología	Personal	Cuestionario	Unidades de contenido Items	Declarado

Nuestro interés es evaluar el significado personal que los estudiantes de psicología alcanzan sobre la probabilidad condicional tras un proceso "estándar" de instrucción, pero esta evaluación ha de tener como pauta el significado institucional en la correspondiente institución de enseñanza (estadística en psicología). No olvidamos que este significado es sólo una parte del sistema de prácticas más complejo y global – el significado global u holístico del concepto, del que aquí se aborda tan sólo una parte y que en este trabajo se considera como dado, no es objeto de investigación.

Por otro lado, el instrumento es elaborado desde una institución diferente, que podemos denominar como *investigación en educación estadística*; en esta institución se comparte un cierto significado del objeto probabilidad condicional, sobre qué sería una evaluación objetiva y los criterios para construir un instrumento aceptable (por ejemplo referidos a fiabilidad, validez, etc.). El investigador es un sujeto de dicha institución; por tanto puede introducir elementos subjetivos tanto en la elaboración del instrumento,

como en la interpretación de resultados, elección de muestra de estudiantes, etc. Es aquí donde se usa el juicio de expertos – a su vez una práctica dentro de dicha institución- como control de la objetividad de proceso. Tanto esta práctica como otras – la serie de análisis psicométricos efectuados- tienen como finalidad evaluar o investigar el significado personal de una forma objetiva.

5.2. PROCESOS DE MUESTREO IMPLICADOS

Para poder comprender la diferencia entre los significados que son objetos del estudio y lo que realmente es posible determinar en la investigación, es importante también reflexionar sobre todo el proceso de inferencia llevado a cabo, desde la definición operacional de la variable hasta la interpretación dada a las respuestas de los estudiantes. Ello nos permite diferenciar entre lo que queremos evaluar y lo que podemos evaluar, así como hasta qué punto podemos generalizar los resultados de un estudio de evaluación y definir en consecuencia algunos criterios que permitan mejorar la generalizabilidad.

Usualmente en la investigación didáctica somos conscientes de que los alumnos participantes en el estudio constituye una muestra de una población real o potencial, a la que queremos extender nuestras conclusiones. Sin embargo, muestreamos también las tareas, contextos, etc. En concreto en el estudio analizado podemos identificar los siguientes procesos de muestreo:

- El investigador está interesado en determinar el significado institucional para poder precisar su variable: qué se entiende como conocimiento de la probabilidad condicional en una cierta institución de enseñanza. Pero no es posible acceder a las clases que se dan en esta institución (que además podrían potencialmente variar de un curso a otro, incluso para un mismo profesor). Una forma de acercarse al significado de interés es el análisis de libros de texto recomendados, que ni siquiera es completo, pues hay un muestreo de Universidades y de libros. Es innegable que el resultado del estudio es un significado diferente, que sería el significado de referencia del proceso de evaluación, en cuanto en base a él se organiza la construcción del cuestionario e interpretación de las respuestas de los alumnos.
- Una vez fijado el significado de referencia hay infinitos posibles instrumentos de evaluación; incluso infinitos posibles cuestionarios. Podemos variar el tipo y número de ítems, el nivel de formalización de los mismos, su dificultad, su contexto. El investigador trata de construir un instrumento lo más objetivo posible, y para ello recopila diferentes ítems, tomados de investigaciones previas en las que han sido evaluados y han dado resultados probados, para cada una de las unidades de contenido. Pero un investigador puede involuntariamente introducir elementos subjetivos en la elección de los ítems o tareas. El recurso a juicio de expertos trata de crear un *significado compartido* de lo que sería un instrumento de evaluación. Las pruebas de los ítems con estudiantes tratan de asegurar la legibilidad y la idoneidad cognitiva de los mismos. El cuestionario resultante define un significado nuevo, sería el significado evaluado, diferente de los anteriores.
- Finalmente el instrumento se prueba con una muestra de estudiantes. La respuesta que cada uno de ellos proporciona a cada ítem no es la única que puede dar. Dependiendo del interés, cansancio, concentración y otros factores, sus respuestas reflejan una parte de lo que el estudiante realmente conoce. Sería el significado declarado que es el finalmente accesible al investigador.

5.3. IDONEIDAD DE UN CUESTIONARIO DE EVALUACIÓN

Pero el interés del estudio no se limita a este significado declarado. El investigador estaría interesado en el significado personal de los alumnos respecto a las tareas propuestas (las respuestas dadas se suponen una muestra representativa de las que darían los mismos estudiantes en la misma prueba en otras ocasiones).

El estudio significativo de los objetos matemáticos debe poner en juego una muestra representativa de las prácticas que constituyen el significado sistémico de los mismos en el seno de un contexto institucional dado (Godino, 1999). Más aún, si las tareas son suficientemente representativas (para evaluar las unidades de contenido definidas), podríamos hacer una inferencia sobre lo que cada alumno de la muestra sería capaz de hacer y decir en otras tareas relacionadas con el concepto. Si las unidades del contenido están bien definidas y representan el concepto de probabilidad condicional, entonces podríamos acercarnos al significado personal de los alumnos de la muestra sobre la probabilidad condicional. Finalmente, los alumnos particulares son una muestra (que suponemos representativa) de otros estudiantes de psicología. Mientras que un profesor se interesa sólo por los alumnos a su cargo, el investigador aspira a obtener conocimiento generalizable sobre las dificultades y capacidades de los estudiantes.

Es claro que la posibilidad de generalizar en cada uno de los pasos descritos depende de la representatividad y la variabilidad de la muestra elegida en cada uno de los procesos de muestreo. Aunque la tarea de conseguir una generalizabilidad completa parece imposible, la investigación didáctica debe aspirar a dar criterios que permitan la construcción adecuada de instrumentos de evaluación o de reinterpretar los criterios clásicos en psicometría. En este sentido, pensamos que podríamos aplicar o extender el concepto de *idoneidad* y sus tipos (Godino, 2003; Godino, Cotreras y Fonts, en revisión) al caso de la evaluación, en el siguiente sentido:

- La *dificultad* de un ítem o tarea daría una medida de su *idoneidad cognitiva*; es decir del grado de representatividad de los significados evaluados respecto a los significados personales.
- La *discriminación* de un ítem valoraría su *idoneidad evaluadora,* un item puede ser adecuado cognitivamente, pero no diferenciar (por ser demasiado fácil) los alumnos que tienen un mayor o menor conocimiento del concepto. Esta idoneidad podría ser un componente de la idoneidad *instruccional*, en cuanto uno de los objetivos de la instrucción es la función evaluadora.
- La *validez de contenido* de un cuestionario indicaría una idoneidad *epistémica*, o grado de representatividad del instrumento en cuanto al significado objeto de evaluación.
- La *fiabilidad* o *generalizabilidad a otros ítems* daría una medida de la estabilidad de la respuesta, es decir sería otro componente de la *idoneidad evaluadora*
- La *validez externa* y *generalizabilidad a otros estudiantes,* sugeriría una *idoneidad generalizadora o externa* en cuanto los resultados se generalizarían a otros estudiantes.

Como conclusión señalamos que el marco teórico nos ha permitido analizar y reflexionar sobre un proceso de investigación – incluso realizado desde una perspectiva muy diferente como es la psicométrica- y reinterpretar desde nuestra perspectiva algunas de sus prácticas y conceptos. El enfoque de la TSF a priori no presupone una metodología única de investigación, sino que puede beneficiarse de múltiples perspectivas a las que a su vez puede enriquecer introduciendo algunos de sus conceptos.

Reconocimientos:
Este trabajo es parte del Proyecto SEJ2004-00789 y Beca FPU: AP2003-5130.

REFERENCIAS

Bar – Hillel, M. (1987). The base rate fallacy controversy. En R. W. Scholz (Ed.), *Decision making under uncertainty.* (pp 39 – 61) Amsterdam: North Holland.

Batanero, C., Estepa, A., Godino, J. y Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27 (2), 151– 169.

Bernard, J. M. (1998). Bayesian inference for categorised data. En H. Rouanet et al. (Eds.), *New ways in statistical methodology* (pp. 159 – 226). Berna: Peter Lang.

Carmines, E. G. y Zeller, R. A. (1979). *Reliability and validity assesment.* Sage University Paper.

Díaz, C. (2004). *Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar.* Trabajo de Investigación Tutelada. Universidad de Granada.

Díaz, C., de la Fuente, I, y Martínez Arias, R. (En prensa). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Epsilon.*

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. En D. Kahneman, P. Slovic y Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Einhorn, H.J. y Hogart, R.M. (1986). Judging probable cause. *Psychological Bulletin.* 99, 3 –19.

Falk, R. (1986). Conditional Probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics.* (pp. 292 – 297). Victoria, Canada: International Statistical Institute.

Feldt, L. S. y Brennan, R. L. (1991). Reliability. En R. Linn (Ed.), *Educational measurement* (pp. 105-146). Nueva York: McMillan.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle lingüistic factors. *Psychological Research,* 50, 123-129.

Gras, R. y Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle *Recherches en Didactique des Mathématiques,* 15(1), 49-95.

Ghiglione, R. y Matalón, B. (1991). *Les enquêtes sociologiques. Théories et practique.* París: Armand Colin.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). En G. Wright y P. Ayton (Eds.), *Subjective probability* (pp. 129-161). Chichester: Wiley.

Godino, J. D. (1999). Implicaciones metodológicas de un enfoque semiótico-antropológico para la investigación en didáctica de la matemática. (Ponencia invitada). En T. Ortega (Ed.), *Actas del III Simposio de la Sociedad Española de Investigación en Educación Matemática* (pp. 196-212). Universidad de Valladolid.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactiques des Mathematiques*, 22, 2/3, 237-284.

Godino, J. D. (2003). *Teoría de las funciones semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática*. Granada: El autor.

Godino, J. D. y Batanero, C. (2003). Semiotic functions in teaching and learning mathematics. En M, Anderson, A. Sáenz-Ludlow, S. Zellweger y V,V, Cifarelli (Eds.), *Educational perspectives on mathematics as semiosis: From thinking to interpreting to knowing* (pp. 149-168). New York: LEGAS.

Godino, J. D., Contreras, A. y Fonts, V. (En revisión). Análisis de procesos de instrucción basado en el enfoque ontológico - semiótico de la cognición matemática. *Recherches en Didáctique des Mathematiques.*

Kelly, I. W. y Zwiers, F. W. (1986). Mutually exclusive and independence: Unravelling basic misconceptions in probability theory. *Teaching Statistics* 8, 96-100.

Lecoutre, B. (1996). *Traitement statistique des données experimentales.* París: CISIA.

León, O. G. y Montero, I. (2002). *Métodos de investigación en psicología y educación.* Madrid: McGraw-Hill.

Martignon, L. y Wassner, C. (2002). Teaching decision making and statistical thinking with natural frequencies. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics.* Ciudad del Cabo: IASE. CD ROM.

Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos.* Madrid: Síntesis.

Maury, S. (1985). Influence de la question dans una épreuve relative á la notion d'independance. *Educational Studies in Mathematics,* 16, 283-301.

Maury, S. (1986). *Contribution à l'étude didacique de quelques notions de probabilité et de combinatoire à travers la résolution de problémes.* Tesis doctoral. Universidad de Montpéllier II.

Millman, J. y Greene, J. (1989). The specification and development of test of achievement and ability. En R. L. Linn (Ed.), *Educational Measurement* (pp. 335 – 366). London*:* Macmillan.

Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO,* 5, 37-55.

Osterlind, S. J. (1989). *Constructing test items.* Boston: Kluwer.

Pollatsek, A., Well, A. D., Konold, C. y Hardiman, P. (1987). Understanding Conditional Probabilities. *Organitation, Behavior and Human Decision Processes.* 40, 255 – 269.

Sánchez, E. (1996). Dificultades en la comprensión del concepto de eventos independientes. En F. Hitt (Ed.), *Investigaciones en Educación Matemática* (pp. 389-404). México.

Sedlmeier, P. (1999). *Improving statistical reasoning. Theoretical models and practical implications.* Mahwah, NJ: Erlbaum.

Serrano Angulo. J. (2003). *Iniciación a la estadística bayesiana.* Madrid: La Muralla.

Teigen, K. H., Brun, W. y Frydenlund, R. (1999). Judgments of risk and probability: the role of frequentistic information. *Journal of Behavioral Decision Making*, *12(2),* 123.

Thorndike, R. L. (1989). *Psicometría aplicada.* Mexico: Limusa.

Totohasina, A. (1992). *Méthode implicative en analyse de données et application á l'analyse de conceptions d'étudiants sur la notion de probabilité conditionnelle.* Tesis Doctoral. Universidad Rennes I.

Tversky, A. y Kahneman, D. (1982a). Causal schemas in judgment under uncertainty. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117-128). Cambridge, MA: Cambridge University Press.

Tversky, A. y Kahneman, D. (1982b). On the psychology of prediction. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 69-83). Cambridge, MA: Cambridge University Press.

Weber, R. P. (1985). *Basic content analysis.* Londres: Sage.