
Cuestiones metodológicas en la evaluación de los conocimientos matemáticos de los alumnos y de su evolución

Carmen Batanero
Universidad de Granada

Introducción

Gran parte de la problemática de investigación didáctica puede formularse por medio de las siguientes preguntas: ¿Cuál es la estructura del conocimiento de los sujetos sobre los distintos campos conceptuales matemáticos? ¿Cuáles son las variables que caracterizan sus estados y posibilitan su control? ¿Cómo evoluciona esta estructura, bien espontáneamente, o por la acción específica de determinadas intervenciones educativas? ¿Qué restricciones contextuales dificultan o condicionan el cambio?

Para responder a estas cuestiones, se requieren instrumentos adecuados, que permitan valorar las matemáticas que conocen y son capaces de usar los alumnos. El diseño de estos instrumentos no es simple, ya que deben tener en cuenta

las situaciones de evaluación, las respuestas, el análisis e interpretación, así como el conocimiento que se evalúa, las características del individuo o grupo que tiene que responder y el propósito de la evaluación (Webb, 1992, p. 668).

En este trabajo nos interesamos en la evaluación de los conocimientos matemáticos, entendida como “el informe comprensivo sobre un sujeto o grupo respecto de las matemáticas o en la aplicación de las matemáticas” (Webb, 1992, p. 662). Aunque

el estudio se abordará desde una perspectiva metodológica, creemos necesario explicitar previamente nuestra posición epistemológica acerca de la naturaleza compleja de las relaciones de los sujetos a los objetos matemáticos, esto es, del conocimiento subjetivo. Asumimos como propia la pregunta de Wheeler (1993), “¿Cómo podemos evaluar lo que no conocemos?” (p. 87), pero también reconocemos la complejidad de la respuesta:

La noción de conocimiento nos parece una y evidente. Pero, en el momento en que se le interroga, estalla, se diversifica, se multiplica en nociones innumerables, planteando cada una de ellas una nueva interrogante (Morin, 1986, p. 18).

Consideramos que el sistema cognitivo de un sujeto (su conocimiento conceptual y procedimental, concepciones, intuiciones, representaciones, esquemas, ...), o sea la trama de objetos personales (Godino y Batanero, 1994) construida en un momento dado, es una totalidad organizada y compleja. Estos objetos personales o mentales son conceptualizados en el trabajo citado como emergentes de los sistemas de prácticas realizadas por una persona ante determinadas situaciones problemáticas o disposiciones del entorno.

La evaluación del conocimiento subjetivo, conceptualizado desde este paradigma de complejidad sistémica, lleva a la necesidad de desarrollar un nuevo marco teórico para la “medición” educacional y psicológica, que supere las limitaciones del enfoque psicométrico clásico. Se precisan, además, realizar procesos de inferencia a partir de las respuestas observadas en situaciones de evaluación, cuya validez y fiabilidad hay que garantizar. La complejidad de esta inferencia se deduce del hecho de que no siempre es posible aplicar a la evaluación de los conocimientos las teorías clásicas psicométricas de maestría de dominio o del rasgo latente (Webb, 1992; Snow y Lohman, 1991).

En los apartados que siguen analizamos con más detalle las cuestiones señaladas, desde una perspectiva metodológica. Usaremos como ejemplo la investigación descrita en Estepa (1994) y Batanero y cols. (1996) en la cual se estudian las concepciones previas de los estudiantes sobre la asociación estadística. Para la determinación de las mismas se utilizó las respuestas de una muestra de 213 alumnos a un cuestionario constituido por 10 problemas. En cada uno de los mismos, el alumno debía emitir un juicio sobre la existencia o no de asociación entre dos variables estadísticas, razonando su respuesta. Como indicadores empíricos de las concepciones de los alumnos, se emplearon el juicio de asociación emitido; la estrategia usada en la solución del problema y las teorías previas expresadas sobre la asociación en las variables del problema. Cada uno de estos aspectos fué tomado

como variable a analizar en la investigación.

Estructura de los conocimientos y su relación con las variables didácticas

La modelización como sistema de los procesos didácticos, ampliamente adoptada por las ciencias sociales quiere enfatizar un punto de vista holístico en el estudio de estos procesos y caracterizar su estructura relacional o principio organizativo, tomando conciencia de la complejidad de los mismos.

El término *estructura* se deriva del análisis de los sistemas, definidos como “unidad global organizada de interrelaciones entre elementos, acciones o individuos” (Morin, 1977, p. 124). Un modelo general de un sistema concreto en un tiempo *t* comprende su composición (conjunto de sus elementos), su entorno (otros sistemas que interaccionan con él) y su estructura (conjunto de relaciones internas y externas de los elementos del sistema). Herman (1990) describe diferentes aproximaciones al estudio de los sistemas: la aproximación atomista se ocuparía sólo del estudio aislado de cada elemento del sistema; la estructuralista privilegia el análisis de la estructura; la causalista o “modelo tipo caja negra” hace abstracción de la estructura del sistema salvo las entradas y salidas del mismo; por último, la sistémica integral es una aproximación que desea estudiar el conjunto de aspectos del sistema completo.

Como primer paso de un enfoque atomista o causalista en el estudio de los conocimientos de los alumnos a un enfoque sistémico, sería preciso abordar un enfoque estructuralista, que podría ser útil en la explicación de una “estática de conocimientos” — la comprensión de los conocimientos en un instante, situación y contexto dado. El estudio de la variación en la composición y estructura de tal sistema, en función de las variables relevantes, conduciría gradualmente al enfoque sistémico:

La estructura de un sistema se concibe como una distribución estadística de los efectos que se obtienen en el funcionamiento general del sistema, mediante la supresión sucesiva de cada elemento. (...) El análisis estructural trata de demostrar cómo las reglas de un determinado fenómeno se implican unas a otras, formando un todo coherente y cómo el cambio de una de las reglas modifica el conjunto general (Cornejo, 1988, p. 63).

En el proceso de análisis estructural se distinguen varias fases:

- La reducción de la variabilidad de lo cualitativo determinando las unidades de análisis. En el trabajo sobre la asociación estadística, la unidad de análisis en el

estudio de las concepciones previas fué el alumno y, dentro de ella el ítem. Para cada uno de los 10 ítems propuestos, se identificaron las variables que hemos descrito con anterioridad.

- La caracterización del conjunto de objetos abstractos, describiendo y construyendo las relaciones que pueden establecerse entre el todo y sus partes, que lleva a la categorización de las diferentes variables. Por ejemplo, las estrategias de resolución de cada uno de los problemas de asociación fueron categorizadas mediante un proceso de diferenciación y comparación sucesivos, en función de los conceptos y teoremas en acto (Vergnaud, 1990) subyacentes en las mismas.
- El establecimiento de un sistema de relaciones entre categorías de la misma o diferentes variables, que pueden ser de dos tipos: relaciones de semejanza (o equivalencia) y relaciones de dominancia (implicación u orden). En el ejemplo citado, el interés se centró en las relaciones de equivalencia entre estrategias, considerando dos estrategias como equivalentes si son intercambiables en el mismo tipo de problemas.

En resumen, el estudio de la interacción en los sistemas complejos nos lleva a la determinación de la estructura relacional entre sus elementos. A su vez, hablar de relación entre dos o más variables, supone plantear el problema de la causalidad. En las ciencias sociales, la noción estricta de causalidad es una hipótesis excesiva, por lo que es preferible aplicar el concepto de orden causal débil, empleado en las técnicas estadísticas de análisis causal, como los modelos de estructuras covariantes. Davis (1985) define la relación de orden causal (débil) entre dos variables del modo siguiente:

X es una causa (débil) de *Y* cuando cada cambio en *X* produce (en promedio) un cambio en *Y* o una constancia en *X* tiende a coincidir con unos valores fijos de *Y* (en promedio).

Esta definición implica, por un lado, la posibilidad de existencia de más de una causa (débil) para una variable dada, y la no coincidencia, en general, de la correlación con la causación. Se admiten relaciones causales (débiles) unidireccionales o recíprocas entre variables. Por otro lado, la definición tiene naturaleza probabilística, por lo que se esperan excepciones a la regla.

Como se ha indicado, los conocimientos de un sujeto sobre un objeto matemático deben ser considerados como un sistema complejo, en el que es posible identificar numerosas variables. Por ello creemos que el estudio de las relaciones de implicación entre las diferentes variables que influyen sobre las respuestas explicitadas en la resolución de una tarea y entre las respuestas a las diferentes tareas puede ser

adecuadamente modelizado mediante una estructura de orden causal (débil) entre las variables intervinientes, entre las que deben figurar, no sólo las derivadas de los problemas (tareas propuestas), sino también las variables del sujeto, del contexto o situación didáctica.

La evaluación del conocimiento como problema de medición

Como consecuencia del enfoque sistémico sobre el conocimiento matemático, la evaluación de los conocimientos de los sujetos requiere la elaboración de un informe global sobre los distintos componentes que permita la toma de decisiones instruccionales o dar respuesta a las cuestiones de investigación planteadas. Para Romberg y cols. (1991):

la intención es evaluar la creación de conocimientos y los procesos implicados, más bien que medir el grado en que los estudiantes han adquirido una parcela del campo de las matemáticas; se precisa una variedad más amplia de medidas, muchas de ellas cualitativas. Se requiere un esfuerzo considerable para encontrar instrumentos adecuados para este propósito (p. 35).

Pero el conocimiento subjetivo tiene el carácter de constructo inobservable, debiéndose emplear distintos indicadores empíricos para su determinación. En general, estos indicadores toman la forma de respuestas (estrategias, argumentaciones, errores y soluciones) a unas tareas prototípicas en las cuales este conocimiento se pone en juego. Estas tareas, o situaciones de evaluación, pueden estar constituidas por una colección de ítems (tests), un problema o proyecto cuya realización puede requerir desde unos minutos hasta varios días, las tareas diarias realizadas por el estudiante en clase o en casa, etc. Además, una respuesta obtenida con fines de investigación no tiene porqué ser registrada sobre un instrumento de evaluación formal, sino que puede estar constituida por la observación cuidadosa del trabajo de los estudiantes durante el desarrollo normal de la instrucción. Por ejemplo, en la investigación de Batanero y cols. (1995), aunque la evaluación de las concepciones previas de los alumnos se llevó a cabo a partir de las respuestas a un cuestionario escrito, el estudio del proceso de aprendizaje de una pareja de alumnos, a partir de la observación de su trabajo en clase, y del análisis de su interacción con el ordenador, permitió indentificar actos de comprensión (Sierpinska, 1994) referidos al concepto de asociación estadística.

Esta variedad de fuentes de recogida de datos y formatos de respuesta plantea el

problema del análisis de las mismas más allá de una simple puntuación sumativa. Un paso esencial en este proceso es la elección de la unidad de análisis y los criterios de categorización de los distintos tipos de respuestas.

La evaluación como medición

En cualquier caso, en el proceso de evaluación de los conocimientos matemáticos de los estudiantes, se deben realizar inferencias sobre constructos inobservables a partir de un conjunto de indicadores empíricos, esto es, se plantea un problema de medición (Sax, 1989; Dane, 1990). Las tareas o items del instrumento de evaluación constituyen una muestra del universo de posibles situaciones problemáticas características del objeto matemático correspondiente. En el caso de la asociación estadística es posible imaginar un sinnúmero de posibles problemas, obtenidos al variar sistemáticamente las variables de tarea identificadas en la investigación psicológica sobre los juicios de asociación. Entre estas variables, podemos citar:

- el tipo de variable estadística cuya asociación se estudia (dos variables cualitativas, lo que daría lugar a una tabla de contingencia; dos variables cuantitativas, lo que llevaría al estudio de la correlación o regresión a partir de un diagrama de dispersión; y una variable de cada tipo, lo que originaría un problema de comparación de dos muestras);
- la intensidad de la asociación y, si es pertinente, el signo; en el caso de que ambas variables sean numéricas, se tendrá en cuenta si la relación entre las mismas es o no lineal;
- si se diferencia claramente entre variables dependiente e independiente; esto es, si la relación es simétrica o asimétrica;
- el motivo por el cual aparece la covariación, que incluye la dependencia de tipo causal, la interdependencia, la concordancia, la dependencia de una tercera variable y la asociación espúrea;
- si la asociación presentada en los datos coincide o no con las teorías previas de los alumnos sobre la misma.

Por otro lado, las respuestas explicitadas al resolver las tareas propuestas son un subsistema del sistema de respuestas que el sujeto considera significativas para resolver la población de situaciones características. Hay un doble proceso de inferencia en el estudio del conocimiento subjetivo: el primero de la muestra de respuestas de los sujetos en las tareas concretas propuestas en el cuestionario, al conjunto de todas las respuestas que los alumnos de la muestra serían capaces de dar, si variásemos los items concretos empleados por otros similares, cambiando los valores dados a las variables de tarea. Esta población de respuestas sería, al menos

teóricamente, potencialmente observable, pero aún queremos efectuar otro proceso de inferencia que es el que lleva de estas respuestas al conocimiento subjetivo, de carácter inobservable.

El interés de la psicometría está en la inferencia para cada uno de los sujetos particulares que han realizado la prueba. Como indica Cuadras (1991, p. 129),

ésta es la base de la inferencia psicométrica, que considera las variables como muestra y los individuos como población, mientras la inferencia estadística considera las variables como población y los individuos como muestra.

Sin embargo, en la investigación didáctica se trata de generalizar tanto respecto a otras tareas como respecto a otros sujetos, por lo que se persiguen los fines de la inferencia psicométrica y estadística conjuntamente.

Como consecuencia, en el análisis de las respuestas e interpretación de los resultados, es preciso tener en cuenta las características esenciales que postulamos para el sistema de indicadores empíricos de los conocimientos matemáticos de los alumnos:

(1) *Multidimensionalidad*: ya que el conocimiento subjetivo sobre un objeto matemático tiene un carácter sistémico, por lo que no podemos asumir a priori, y en general, la existencia de un rasgo latente cuantificable que indique grados de adquisición del mismo; por el contrario, el conocimiento estaría constituido por aspectos múltiples, cuya construcción precisaría la realización de distintos “actos de comprensión” (Sierpiska, 1994). Metafóricamente, el conocimiento se asemeja más a un árbol que a una escalera o rampa ascendente.

Así, respecto a la asociación estadística, los tres tipos básicos de tareas respecto a las cuales es posible pedir un juicio de asociación: nubes de puntos, comparación de muestras y tablas de contingencia, requieren muy diversas competencias para su resolución. Por ejemplo, es posible juzgar la asociación en una nube de puntos, empleando propiedades del cuadro gráfico, como el crecimiento o la dispersión de la nube. Por el contrario, en la tabla de contingencia la asociación no se juzga a partir de los valores de las variables, sino de la comparación de las frecuencias de los mismos.

(2) *Carácter cualitativo*: que se manifiesta en cada uno de estos componentes; así para una propiedad dada no pueden reducirse las posibles manifestaciones que los alumnos hacen de la misma a “conocerla” o “no conocerla”; interesa diferenciar los distintos tipos de errores y estrategias, que, en general, no pueden ser ordenados en una escala numérica. Así, todas las estrategias erróneas en el estudio de la asociación en una tabla de contingencia no son equivalentes. Si el alumno compara todas las

distribuciones condicionales, pero emplea frecuencias absolutas, la estrategia es errónea, pero el alumno ha comprendido que el problema implica la comparación de dos probabilidades. Si, por el contrario, utiliza sólo una distribución condicional, ello es indicativo de una concepción localista (definida por Estepa, 1994), ya que basa su juicio en una parte de los datos, sin tener en cuenta la representatividad de estos respecto al conjunto de datos completo.

(3) *Importancia de las interacciones*: puesto que nos encontramos en un sistema, este no queda determinado por la enumeración de sus componentes. La estructura del mismo, esto es, las interacciones entre los diferentes componentes del significado, es un aspecto esencial, porque nos revela posibles variables sobre las que actuar si queremos lograr una evolución de estos conocimientos

Técnicas de análisis e interpretación de datos

Como afirma Cornejo (1988),

El mayor problema que se ha planteado hasta fechas muy recientes al estudio simultáneo de un gran número de variables ha sido la enorme dificultad de captar el conjunto sin perder la red fina de interrelaciones específicas (p. 6).

Una posible solución al estudio de la complejidad de los sistemas didácticos ha sido reduccionista en un doble sentido: por un lado, se dicotomizan las variables dependientes cualitativas que hemos señalado, reduciendo las respuestas a los diferentes ítems a las posibilidades de “correcta” o “incorrecta” — o a una graduación en unas pocas modalidades ordenadas—. Por otro lado, se ha transformado el conjunto de variables dependientes en una única variable numérica derivada: “número de respuestas acertadas” o “puntuación total obtenida”. A partir de esta puntuación se han construido modelos explicativos en función de posibles variables del sujeto, de contexto y de instrucción. Una de las razones para el empleo de esta estrategia simplificadora es la limitación de la tecnología estadística utilizada, básicamente modelos paramétricos lineales.

Pero si se quiere profundizar en el estudio de la estructura de los sistemas cognitivos de los alumnos y en el de su evolución como consecuencia de la interacción con el medio instruccional, se debería tratar globalmente el conjunto formado por cada una de las respuestas de un mismo sujeto y, en los casos en que se considere conveniente, respetar su carácter cualitativo. Esto requiere nuevas técnicas de análisis de los datos, como el análisis de correspondencias, métodos de

clasificación, etc. De acuerdo con Gras (1992), el análisis multivariante constituye una ruptura epistemológica con la estadística clásica que concierne a la vez a los objetivos, los medios técnicos para alcanzarlos, el tipo de datos tratados (número, naturaleza, variedad), los sujetos de análisis (variables o individuos) y el proceso (de los datos al modelo y no a la inversa), los métodos matemáticos empleados y los conceptos implícitos en los mismos.

Estas nuevas técnicas, junto con el análisis teórico de los fenómenos bajo estudio, pueden poner de manifiesto relaciones de causalidad débil entre el conjunto de variables. Esto supone un ensanchamiento de los métodos psicométricos tradicionales y puede permitir una mejora tanto en la validez ecológica de las investigaciones didácticas — esto es, su aplicación a casos naturales concretos—, como en la comprensión del funcionamiento de los sistemas didácticos. Los nuevos modelos de medición educacional y psicológica son también apoyados por los desarrollos actuales en el campo de la psicología cognitiva (Snow y Lohman, 1991).

En síntesis, una vez elegida una muestra representativa de problemas-tipo característicos del objeto conceptual de interés, se analizarían las respuestas de los alumnos junto con las variables incluídas en los problemas, con objeto de determinar la ordenación de dichas variables en cuanto a la variabilidad de las respuestas que producen en los alumnos. El conjunto de respuestas de los sujetos a los ítems se toman como coordenadas de dos sistemas de puntos en un doble espacio vectorial multidimensional empírico (espacio de ítems/espacio dual de sujetos). La metáfora geométrica empleada para modelizar el conjunto de datos tiene una gran potencia, pues permite, mediante la definición de una distancia entre puntos (sujetos o variables) dotar de una topología a estos espacios, a los que es posible aplicar técnicas topológicas, algebraicas y geométricas.

Un primer grupo de técnicas de clasificación (Benzecri, 1994; Lerman, 1981; Celeux y cols., 1989) permiten agrupar individuos y variables mediante algoritmos de tipo muy diverso (análisis “cluster” de sujetos o variables; jerárquico o no jerárquico; ascendente o descendente). La proximidad entre ítems señalará la semejanza de respuestas respecto a estos ítems por parte de los sujetos, o lo que es lo mismo, la presencia de similitudes en componentes del conocimiento relacionadas con la resolución de estos ítems (análoga interpretación tiene la proximidad entre sujetos). En la investigación de Estepa (1994) el análisis “cluster” jerárquico de las respuestas (dicotomizadas en correctas o incorrectas) a los diversos ítems del cuestionario empleado, ha permitido identificar agrupaciones de ítems que indican la presencia de factores diferenciados en las competencias necesitadas para resolver los problemas propuestos.

Por otro lado, las técnicas factoriales como el análisis factorial, escalamiento multidimensional, análisis de correspondencias simple y múltiple, (Benzecri, 1994; Pagés y Escofier 1995) utilizan las propiedades de los espacios vectoriales euclídeos para determinar las posibilidades de reducción de la dimensión de los datos y facilitar una representación gráfica de los mismos que permita una síntesis global. El hecho de que este sistema complejo pueda o no estructurarse en unas pocas dimensiones básicas, y que las características de la estructura puedan variar por la acción de variables del sujeto o instruccionales puede ser explorado con ayuda de estas técnicas. La posibilidad de proyectar sobre los ejes obtenidos los individuos o las variables suplementarias permite conocer la correlación con los factores de cualquier otra variable de interés, incluso no perteneciente al dominio estudiado (Escofier y Pagés, 1988). El papel de las variables suplementarias consiste en ampliar el contexto de interpretación.

En particular, en el análisis factorial de correspondencias es posible el trabajo con variables cualitativas y sus propiedades de equivalencia distribucional y representación dual lo hacen un instrumento especialmente valioso. La representación — como sujetos suplementarios— de las variables de tarea en el espacio individuos — tareas permitirá formular y contrastar teorías respecto a las variables y valores de las mismas que se relacionan con una respuesta similar a estos conjuntos de items relacionados. Esta posibilidad es considerada como un nuevo enfoque de la regresión y permite, no sólo ver en perspectiva cualquier otra variable de interés, sino posicionar “perfiles ideales” propuestos por la teoría, cuya posición privilegiada en el plano de proyección tiene siempre un carácter de “prueba” (Pagés y Escofier, 1995). El estudio de los factores principales en relación con estas variables de tarea permitirá identificar relaciones de dependencia — interacción entre estas variables y clasificarlas en grupos de importancia decreciente en cuanto a su efecto sobre la variabilidad de las respuestas.

En la investigación sobre la asociación estadística, el análisis de correspondencias de las estrategias empleadas en los juicios de asociación en tablas de contingencia reveló dos factores que explicaban conjuntamente el 82.6% de la inercia. El primer factor fué identificado como la dimensionalidad de la tabla y mostró que al aumentar el número de filas y columnas de la tabla, los alumnos empleaban, en general, estrategias más elaboradas. Asimismo ocurría en el caso de la independencia perfecta. Por ello, mientras que con una estrategia intuitiva parcialmente correcta muchos alumnos fueron capaces de proporcionar un juicio correcto de asociación en tablas 2x2, en las que existiese una dependencia marcada; para la apreciación de la independencia, incluso en tablas 2x2 o la dependencia en una tabla de dimensión

superior, las estrategias parcialmente correctas, resultaron ser insuficientes.

El segundo factor mostró el peso de las teorías previas sobre la asociación esperada en el contexto del problema sobre las estrategias del alumno. Cuando estas expectativas no coincidieron con la asociación empírica de los datos del problema, hubo una tendencia a sustituir las estrategias correctas o parcialmente correctas por estrategias incorrectas, como medio de sostener sus teorías, incluso a pesar de la evidencia en contra, manifestándose claramente el mecanismo de la “correlación ilusoria” (Chapman y Chapman, 1969).

Finalmente, citamos la técnica de análisis implicativo, puesta a punto recientemente por el equipo de R. Gras y sus colaboradores. A partir de la noción de cuasi-implicación, medida por una intensidad, y la de grafo de implicación (Gras, 1995), este método permite representar, en el seno de una familia de variables, el orden (preorden) parcial que lo estructura. Mientras que las técnicas de análisis “cluster” y de correspondencias se basan en las nociones de similaridad y distancia, que son simétricas, la implicación entre variables es una relación asimétrica. Finalmente, el conjunto de clases se organiza mediante una estructura arborescente, que recuerda la clasificación jerárquica, aunque el nexo entre dos nudos del árbol es, en este caso, orientado. En nuestra opinión, la aplicación del análisis implicativo al estudio de las relaciones entre el éxito/fracaso en los items de una prueba, puede ampliar con notoria ventaja los estudios clásicos de escalabilidad en cuestionarios, por medio del escalograma de Guttman, ya que éste presupone una ordenación lineal de los items, que es difícil encontrar en la mayor parte de las aplicaciones.

Desde diferentes puntos de vista, las técnicas de clasificación jerárquica, el análisis factorial de correspondencias y el análisis implicativo complementan el estudio cualitativo de la estructura de las respuestas de los alumnos. Habría que añadir el resto de técnicas multivariantes, aunque la extensión de este trabajo no nos permite detenernos en ellas.

Las cuestiones de validez y precisión

El sistema de respuestas a una muestra de tareas constituye una “medida” de los conocimientos del sujeto. Como en todo proceso de muestreo, estamos sujetos a la posibilidad de dos tipos de errores: sistemáticos (sesgos) y aleatorios. La validez de las conclusiones de los estudios dependerá de la falta de sesgo, esto es de la correspondencia entre las inferencias efectuadas a partir de los indicadores empíricos y los constructos teóricos.

Un aspecto a destacar en la evaluación matemática es que ésta se realiza desde el

punto de vista de una institución dada, por lo que esta validez será siempre relativa a dicha institución. Así, no es lo mismo decir que un alumno de Secundaria, que recibe un curso de estadística descriptiva, conoce el concepto de asociación, que decir que un alumno que finaliza la especialidad de estadística, dentro de la Licenciatura de Matemáticas, conoce este concepto. Lo que se considera adecuado para asegurar el dominio del concepto en uno y otro caso es muy diferente, por lo que un instrumento de evaluación del conocimiento sobre la asociación puede ser válido para uno de estos casos y no serlo para el otro.

En la teoría del significado formulada por Godino y Batanero (1994) diríamos que los sistemas de prácticas asociados al campo de problemas de la asociación estadística no son los mismos en las distintas instituciones. Por tanto, para analizar las cuestiones de validez de los instrumentos, y en general del diseño de las investigaciones, referidas a la determinación de los conocimientos y su evolución, nos parece esencial la distinción entre conocimiento subjetivo (personal) e institucional. Las instituciones escolares constituyen, según esta conceptualización, formaciones epistemológicas diferenciadas en las que los saberes “sabios” — considerados habitualmente como el conocimiento objetivo — sufren adaptaciones o transposiciones didácticas (Chevallard, 1991); estos conocimientos institucionales servirán de punto de referencia para la organización de los procesos de instrucción y la fijación de las pautas de evaluación de los conocimientos de los alumnos. En términos de la teoría estadística del muestreo, podemos expresar esta problemática diciendo que las “poblaciones objetivo” que se tratan de alcanzar mediante el estudio inferencial realizado en la evaluación son diferentes.

La construcción de los instrumentos de evaluación, que garanticen la validez de contenido para una finalidad específica, debe partir, por tanto, de un análisis pormenorizado de los significados institucionales de los objetos correspondientes. El análisis de las variables didácticas del campo de problemas proporcionará un criterio para estructurar la población de las posibles tareas de las cuales debe extraerse una muestra representativa, si se quiere garantizar la validez del instrumento de evaluación. Asimismo, la validez debe ser pensada como una variable, no dicotómica, sino componencial y sistémica como corresponde a la naturaleza del conocimiento que se trata de evaluar. Como señala Messick (1991), la validez no es, por otra parte, una propiedad exclusiva del instrumento de recogida de datos, sino de todo el proceso de inferencia. Afecta pues a la falta de sesgo en la muestra de sujetos empleada (validez externa) en el análisis estadístico de datos (validez estadística), al instrumento (validez de construcción) y al control adecuado de las variables intervinientes (validez interna) (Cook y Campbell, 1979).

Minimización del error de muestreo (precisión)

Otro problema de diseño del muestreo de las situaciones de evaluación sería la determinación de las variables de tarea y valores de las mismas que maximicen la varianza de las respuestas de los alumnos entre estratos, con objeto de conseguir, no sólo la representatividad muestral, sino la máxima reducción del error de muestreo. Aunque, como hemos indicado, es posible efectuar un análisis a priori de tales variables, sólo mediante el estudio experimental sobre los sujetos concretos a quienes va dirigida la prueba podremos confirmar las hipótesis generadas en el análisis a priori, relativas a la variabilidad de los tipos de respuestas.

Una vez conocidas estas varianzas, aplicando los principios del muestreo estratificado sería posible determinar la forma óptima de selección de una muestra, de modo que se consiguiese la mayor eficiencia — para la evaluación de conocimientos — con un mismo tamaño de muestra. Al determinar la prioridad (en el sentido de maximización de la variabilidad de respuestas) de las variables de tarea, para una situación o un grupo de situaciones características de un objeto, se proporcionarían criterios para el refinamiento progresivo de los instrumentos de evaluación.

La *fiabilidad* o precisión de las inferencias dependerá de la magnitud del error aleatorio (Feldt y Brennan, 1991), que puede ser debido a diferentes facetas o componentes: variabilidad entre sujetos, variabilidad entre preguntas, etc. Por tanto, consideramos que el marco más adecuado para el análisis de estas facetas lo ofrece la teoría de la generalizabilidad (Brenan, 1983). Hemos de tener en cuenta, también, que, puesto que nos hallamos en un problema de muestreo estratificado, es lógico esperar una gran variabilidad entre los diferentes estratos, ya que el diseño del cuestionario ha ido buscando este criterio de máxima variabilidad. Esto lleva implícito, como contrapartida, la homogeneidad dentro de los estratos, lo que nos puede autorizar, en caso necesario, a preferir un menor tamaño de muestra para cada uno de los estratos particulares, y un mayor número de estratos. El análisis global de todos estos aspectos es el que nos permitirá deducir criterios o medidas de la generalizabilidad de los resultados.

En el caso del trabajo de Estepa (1994), aplicando la teoría de la generalizabilidad se han obtenidos dos coeficientes diferenciados de generalizabilidad, a partir del análisis de varianza de medidas repetidas de los porcentajes de acierto en los ítems, en función de los factores alumno (entre) e ítem (dentro). El primero de ellos, como medida de la generalizabilidad de los resultados, cuando se fijasen los alumnos y se sustituyesen los ítems por otros pertenecientes a la misma población, dió un valor moderado. Ello es lógico, debido a la gran cantidad de variables de tareas presentes en los ítems.

El segundo coeficiente, para el cual se obtuvo un alto valor, mide las posibilidades de extender las inferencias, cuando el mismo cuestionario fuese usado con otros alumnos de Secundaria. En consecuencia, de este estudio vemos que, en general, cuando tratamos con conceptos complejos, las posibilidades de extensión de los estudios será mayor al variar la población de alumnos, que al variar la población de situaciones de evaluación. Ello concuerda con el carácter local de las concepciones identificadas en los estudios didácticos (Artigue, 1990), estrechamente ligadas al saber puesto en juego por las situaciones concretas.

Enfoque experimental y contraste de hipótesis

El enfoque clásico en el estudio del efecto de las variables de tarea e instruccionales sobre las respuestas de los alumnos ha sido el experimental comparativo. Centrándose en un único — o unos pocos — tipo de tareas, se analizan sus variables y se intenta contrastar estadísticamente sus efectos, mediante un experimento, en el que se varían sistemáticamente unas variables, controlando las restantes y comparando los resultados con grupos equivalentes. El control adecuado de todas las variables intervinientes es esencial, debido a la hipótesis de independencia, subyacente en los modelos lineales empleados en este enfoque. El no cumplimiento de esta hipótesis puede invalidar por completo las inferencias realizadas a partir de dichos procedimientos estadísticos (Bollen, 1989). En un extremo opuesto podemos situar los estudios etnográficos interpretativos, basados principalmente en la observación naturalística de casos y, por tanto, en la ausencia de control y manipulación de las situaciones.

En el enfoque sistémico, cuasiexperimental e inductivo que proponemos para los estudios didácticos, en gran medida concordante con la denominada en Francia metodología de ingeniería didáctica (Artigue, 1994), el análisis y control de las variables es también el ideal que nos permitirá obtener la representatividad a que hemos hecho alusión al hablar de los procesos de muestreo implícitos. La metodología de ingeniería didáctica se basa en el estudio de casos y el uso de criterios de validación interna, obtenidos por la confrontación de los análisis a priori de los conocimientos puestos en juego y de las variables de control de la situación experimental diseñada. El investigador planifica los experimentos de enseñanza y decide actuar sobre un cierto número de variables de maniobra de los sistemas didácticos, variables que determinan la organización global y local de la ingeniería.

Las posibilidades de control de las variables intervinientes pueden ser el mantenimiento constante, la aleatorización, la variación controlada o la medición e

inclusión en el análisis. Las consecuencias de cada uno de estos tipos de control, que varían desde una restricción excesiva de la generalizabilidad de los resultados, en el caso de mantenimiento constante, al rápido crecimiento del tamaño del experimento, si se desea medir el efecto de las variables de interés.

Como hemos indicado, el enfoque experimental ha estado asociado al análisis confirmatorio de los datos, basado en el contraste de hipótesis. Gras (1992) señala las siguientes limitaciones del método experimental confirmatorio:

a) El objetivo de este enfoque es la construcción de modelos probabilísticos relativos a las variables observadas. Para ello se ajustan deductivamente a una o varias “variables dependientes” el resto de las variables, con el fin de obtener una predicción y los datos se “ajustan” al modelo. Por el contrario, en un enfoque inductivo, el modelo se deduce a partir de los datos, conjuntamente con la teoría bajo la cual estos datos han sido recogidos. Como señala Chatfield (1988, p. 17) el propósito de la construcción de modelos no debiera ser obtener “el mejor ajuste”, sino construir un modelo consistente, no sólo con los datos, sino con el conocimiento que sirve de base a la investigación y con otros conjuntos de datos semejantes tomados con anterioridad.

b) El contraste de las hipótesis se efectúa mediante el rechazo de las “hipótesis nulas” o de no diferencia. Pero, como expone Seidenfeld (1979), en la teoría del contraste de hipótesis, el rechazo de la hipótesis nula sólo presupone que se rechaza el hecho de que las diferencias observadas sean debidas al azar. La demostración o corroboración de que las diferencias son debidas a la supuesta variable que actúa como efecto no queda garantizada por el contraste de hipótesis. Ello será el fruto de una labor de nuevas replicaciones del experimento y de la aplicación de controles para rechazar las posibles explicaciones alternativas o riesgos de invalidez (Cook y Campbell, 1979). A pesar del reconocimiento universal de la potencia de la aleatorización, los experimentos aleatorizados son muy raros en la investigación social. Ello añade la necesidad de revisar el empleo de las técnicas estadísticas, ya que las técnicas habituales no siempre proporcionan estimaciones apropiadas de los efectos de los tratamientos en los cuasi-experimentos.

Chatfield (1988) discute la distinción entre análisis confirmatorio y exploratorio de datos (EDA) e indica que prácticamente no se encuentra un análisis confirmatorio puro de los datos en la literatura de investigación. La mayoría de los resultados “estadísticamente significativos” publicados provienen de conjuntos de datos singulares — en el sentido de que se recogen una sólo vez para la investigación concreta. Estos conjuntos de datos tienen también naturaleza exploratoria, aunque se sepa lo peligroso que es descubrir una característica de interés en un conjunto de datos y

contrastar esta característica sobre el mismo conjunto de datos.

Este autor prefiere hacer una distinción entre las técnicas iniciales de análisis de datos (IDA) y “análisis definitivo” de los datos. En el IDA incluye tanto el EDA, la estadística descriptiva clásica y las técnicas multivariantes exploratorias, como el empleo informal de métodos inferenciales, en especial el contraste de hipótesis y los modelos lineales. En un contraste de hipótesis, “con tamaños grandes de muestra, los resultados serán casi siempre “significativos”, incluso aunque los efectos sean muy pequeños” (Chatfield, 1988, p. 51).

Además, un resultado significativo no proporciona una “prueba”, sólo evidencia de tipo formal. Cualquier intento de probar la probabilidad de que la hipótesis sea cierta implicará al teorema de Bayes y requerirá la probabilidad a priori de la hipótesis, y, aún así, existe una controversia no resuelta sobre si la probabilidad obtenida puede ser realmente objetiva, o solamente una revisión de nuestro “grado de creencia”, subjetivo y, en consecuencia, falible (Rivadulla, 1991). Un experimento simple debiera formar parte de una cadena de experimentos similares y debiéramos buscar más que el resultado estadísticamente significativo la “similitud significativa” en diferentes experimentos. El IDA (Initial Data Analysis) debiera ser el comienzo obligado para generar hipótesis sensatas sobre los datos, para sugerir cual debería ser el tipo de análisis estadístico definitivo e incluso para sugerir que un contraste de hipótesis es innecesario o inadecuado por alguna de las razones siguientes:

- Puede deducirse claramente a partir del IDA si los resultados concuerdan significativamente con las hipótesis de interés, debido a que los efectos son apreciables sin necesidad de un análisis posterior.
- El IDA puede sugerir que los datos no son apropiados para un contraste formal, debido a la contaminación (existencia de variables que sesgan las posibles conclusiones), a la falta de aleatorización o al incumplimiento del requisito de aplicación de los contrastes estadísticos.

En resumen, creemos que los principios básicos del diseño experimental: control, replicación — y cuando sea posible la aleatorización — deben ser, en lo posible, utilizados en la investigación sobre los conocimientos de los alumnos y su evolución. Proponemos, sin embargo, sustituir los métodos inferenciales clásicos de análisis de datos por un análisis en dos etapas: En primer lugar, un análisis inicial de datos (IDA), potenciando en particular el empleo de técnicas multivariantes y técnicas de análisis de datos cualitativos. Si el resultado obtenido lo permite, deberíamos progresivamente irnos acercando a la construcción de modelos (lineales o no lineales, univariantes o no) empleando el enfoque confirmatorio en sucesivos conjuntos de

datos.

Conclusiones

La perspectiva sistémica en los estudios didácticos, junto con las nuevas tecnologías de análisis de datos disponibles, implican no solo un cambio en los métodos de recogida y transformación de los datos sino en la propia problemática de investigación. Esta se puede enfocar hacia la búsqueda de la estructura relacional en los distintos subsistemas de los sistemas didácticos y hacia la caracterización de su dinámica, siguiendo las pautas de la denominada *ingeniería didáctica* (Artigue, 1994). También implican cambios profundos en el paradigma de evaluación de los conocimientos matemáticos llevada a cabo con fines de investigación didáctica. Particularmente se deben analizar las cuestiones de validez de las investigaciones y la precisión de las mediciones educacionales, teniendo en cuenta la multiplicidad de fuentes de variación e interacción entre variables puestas en juego.

Referencias

- Artigue, M. (1994). Didactical engineering as a framework for the conception of teaching products. En R. Biehler, R. W. Scholz, R. Straesser y Winkelman (Eds). *Didactics of mathematics as a scientific discipline*. Dordrecht: Kluwer.
- Artigue, M. (1990). Epistémologie et didactique. *Recherches en Didactique des Mathématiques*, 10(2-3), 241-286.
- Batanero, C., Estepa, A., Godino, J. D. y Green, D. R. (1996). Judgments of association in contingency tables: An empirical study of students' intuitive strategies and preconceptions. *Journal for Research in Mathematics Education*, 27(2), 151-169.
- Benzecri, J. P. (1994). *La pratique de l'analyse des données*. París: Dunod.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: J. Wiley.
- Brennan, R. L. (1983). *Elements of generalizability theory*. Iowa: ACT Publ.
- Celeux, G., Diday, E., Govart, G. y Le Chevalier, Y. (1989). *Classification automatique des données*. París: Dunod.
- Chapman, L. J. y Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.
- Chatfield, C. (1988). *Problem solving: a statistician's guide*. London: Chapman & Hall.
- Chevallard, Y. (1991). *La transposition didactique. Du savoir savant au savoir enseigné*. Grenoble: La Pensée Sauvage. (2. ed.).
- Cook, T. D. y Campbell, D. T. (1979). *Quasi-experimentation. Design and issues for field setting*. Chicago: Rand Mc Nally.
- Cornejo, J. M. (1988). *Técnicas de investigación social: el análisis de correspondencias*. Barcelona: P.P.U.
- Cuadras, C. M. (1991). *Métodos de análisis multivariante*. Barcelona: P.P.U. (2. Ed.).

- Dane, F. C. (1990). *Research methods*. Pacific Grove, California: Thompson Information Publishing Group.
- Davis, J. A. (1985). *The logic of causal order*. London: Sage University Paper.
- Estepa, A. (1994). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis Doctoral. Universidad de Granada.
- Feldt, L. S. y Brennan, R. L. (1991). Reliability. En R. L. Linn, (Ed.) *Educational measurement* (Third ed.) (pp. 105-146). New York: American Council on Education and Macmillan Publ.
- Gras, R. (1992). L'analyse des données: une méthodologie de traitement de questions de didactique. *Recherches en Didactique des Mathématiques*, 12(1), 59-72.
- Gras, R. (1995). L'analyse statistique implicative: Introduction. En R. Gras (Ed.) *Actes du Colloque "Méthodes d'analyses statistiques multidimensionnelles en Didactique des Mathématiques"* (pp. 129-143). Caen: A.R.D.M.
- Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, 14(3), 325-355.
- Herman, J. (1990). *Analyse de données qualitatives. V. 2: Traitement d'enquêtes, modèles multivariés*. Paris: Masson.
- Lerman, I. C. (1981). *Clasificación et analyse ordinaire des données*. Paris: Dunod.
- Messick, S. (1991). Validity. En R. L. Linn, (Ed.) *Educational measurement* (Third ed.) (pp. 13-104). New York: American Council on Education and Macmillan Publ.
- Morin, E. (1977). *El método I. La naturaleza de la naturaleza*. Madrid: Cátedra, 1986.
- Morin, E. (1986). *El método. El conocimiento del conocimiento*. Madrid: Cátedra, 1988.
- Pages, J. y Escofier, B. (1995). Introduction a l'analyse en composantes principales a partir de l'étude d'un tableau de notes. En R. Gras (Ed.) *Actes du Colloque "Méthodes d'analyses statistiques multidimensionnelles en Didactique des Mathématiques"* (pp. 39-52). Caen: A.R.D.M.
- Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropos.
- Romberg, T. A., Zarinnia, E. A. y Collis, K. F. (1991). A new world view of assessment in mathematics. En G. Kulm, (Ed.) *Assessing higher order thinking in mathematics*. Washington: American Association for the Advancement of Science.
- Sax, G. (1989). *Principles of Educational and psychological measurement and evaluation*. Belmont, California: Wadsworth Publishing Company (3. ed.).
- Seidenfeld, T (1979). *Philosophical problems of statistical inference: Learning from R. A. Fisher*. Dordrech: Reidel.
- Sierpinska, A. (1994). *Understanding in mathematics*. London: Falmer Press.
- Snow, R. E. y Lohman, D. R. (1991). Implication of cognitive psychology for educational measurement. En R. L. Linn, (Ed.) *Educational measurement* (Third ed.) (pp. 263-331). New York: American Council on Education and Macmillan Publ.
- Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques*, 10(2-3), 133-170.
- Webb, N. L. (1992). Assessment of students' knowledge of mathematics: A step toward a theory. En D.A. Grouws, (Ed.) *Handbook of research on mathematics teaching and learning*. New York: Macmillan.
- Wheeler, D. (1993). Epistemological issues and challenges to assessment: What is mathematical knowledge? En M. Niss, (Ed.) *Investigations into assessment in mathematics education; An ICMI Study* (pp. 87-95). Dordrecht: Kluwer.

Carmen Batanero, Departamento de Didáctica de la Matemática, Facultad de Ciencias de la Educación, Universidad de Granada, 18071 Granada, España. Endereço electrónico: cbatanero@goliat.ugr.es.

RESUMEN. En este trabajo analizamos el problema de las inferencias sobre los conocimientos matemáticos de los alumnos, su estructura y evolución, a partir de las respuestas observadas en situaciones de evaluación. Sobre este problema abordamos los puntos siguientes, desde una perspectiva de investigación didáctica: (1) La necesidad de realizar un estudio sistémico, tanto cuantitativo como cualitativo del conjunto de respuestas proporcionadas por cada sujeto a las tareas utilizadas en la evaluación. (2) Las posibilidades ofrecidas por el análisis multivariante de datos para el establecimiento de modelos de causalidad débil entre el sistema de respuestas y las variables de tarea, del sujeto e instruccionales. (3) La problemática asociada a la generalización de inferencias sobre los conocimientos a partir de muestras particulares de alumnos y situaciones de evaluación. (4) Respecto del diseño de las investigaciones didácticas, se comparan el enfoque sistémico cuasi-experimental con el experimental confirmatorio, basado en el contraste de hipótesis, concluyendo en el carácter complementario de ambos.

ABSTRACT. In this paper we are interested in the problem of drawing out inferences on students' mathematical knowledge, its structure and evolution, from the observed answers obtained from assessment situations. About this problem we approach the following points from a research perspective: (1) The necessity to accomplish a systemic study, both quantitative and qualitative, of the set of answers given by each subject to the assessment tasks. (2) The possibilities provided by multivariate data analysis for building models of weak causality among the answers and the task, subject and instructional variables. (3) The issues linked to generalizing inferences on knowledge from samples of students and assessment situations. (4) As regard to educational research, we compare the systemic quasi-experimental and the confirmatory experimental paradigms, pointing out to their complementary nature.