ANÁLISIS DE LA COMPLEJIDAD SEMIÓTICA DE LOS GRÁFICOS PRODUCIDOS POR FUTUROS PROFESORES DE EDUCACIÓN PRIMARIA EN UNA TAREA DE COMPARACIÓN DE DOS VARIABLES ESTADÍSTICAS

Batanero, Carmen[1]; Arteaga, Pedro[1] y Ruiz, Blanca[2]
[1] Departamento de Didáctica de las Matemáticas. Universidad de Granada
[2] Departamento de Didáctica de las Matemáticas. Instituto Tecnológico de Monterrey
batanero@ugr.es
pedroarteagacezon@yahoo.com
bruiz@itesm.mx

Resumen. En este trabajo analizamos los gráficos producidos por 93 futuros profesores de educación primaria en una tarea de comparación de dos variables estadísticas. Se define un nivel de complejidad semiótica que permite clasificar los gráficos construidos y relacionarlos con el nivel de lectura de gráficos y las conclusiones obtenidas. Los resultados muestran que sólo parte de los participantes producen un gráfico con la complejidad necesaria para obtener una conclusión y que una parte de los que lo consiguen no llegan a un nivel suficiente de lectura o bien no logran una conclusión completa sobre la pregunta planteada.

Palabras clave. Gráficos estadísticos, futuros profesores, comparación de variables estadísticas.

Analysing the Semiotic Complexity of Graphs Produced by Future Primary School Teachers in a Task Related to the Comparison of Two Statistical Variables

Summary. In this paper, we analyse the graphs produced by 93 prospective primary school teachers in a task comparing two statistical variables. We define a level of semiotic complexity that serves to classify the graphs produced. We relate this level to the level in reading the graphs and the conclusion obtained. Results show that only some participants produce a graph with enough complexity to get an adequate conclusion and part of them either do not reach an adequate level of reading or do not get a complete conclusion in relation to the research question.

Keywords. Statistical graphs, future teachers, comparing two statistical variables.

PROBLEMA DE INVESTIGACIÓN Y SU RELEVANCIA

El problema de investigación abordado en este trabajo es el estudio de la capacidad de los futuros profesores de educación primaria en la construcción e interpretación de gráficos estadísticos cuando trabajan con un proyecto abierto de análisis de datos. Éste es un problema de gran interés, debido a la inclusión en los recientes Decretos de Enseñanzas Mínimas de los gráficos estadísticos desde el primer ciclo de la Educación Primaria y a la recomendación de que la enseñanza de la estadística en este nivel educativo se haga a través del trabajo con proyectos.

Además, el lenguaje gráfico es esencial en la organización y análisis de datos, al ser un instrumento de *transnumeración,* una forma básica de razonamiento estadístico (Wild y Pfannkuch, 1999) que produce nueva información, al cambiar de un sistema de representación a otro. Por ejemplo, al pasar de una lista de datos desordenada a un histograma, se visualiza la moda y se percibe la simetría o asimetría de la distribución.

La construcción e interpretación de gráficos estadísticos es también parte importante de la cultura estadística que es la unión de dos competencias relacionadas*: «a) Interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos y b) discutir o comunicar opiniones respecto a tales informaciones estadísticas cuando sea relevante»* (Gal, 2002, pp. 2-3).

Una persona culta debería poder leer críticamente los gráficos estadísticos que encuentra en la prensa, Internet, medios de comunicación y trabajo profesional. Esto supone no sólo la lectura literal del gráfico, sino poder identificar las tendencias y variabilidad de los datos, así como detectar los posibles errores conscientes o inconscientes que puedan distorsionar la información representada (Schield, 2006). Asimismo, debería conocer los convenios de construcción de los diferentes tipos de gráficos y ser capaz de construir correctamente un gráfico sencillo.

Esta capacidad es especialmente importante en el caso de futuros profesores de educación primaria. En el Decretos de Enseñanzas Mínimas (MEC, 2006) se incluyen los siguientes contenidos, dentro del Bloque 4, sobre tratamiento de la información, azar y probabilidad:

– *Primer ciclo:* Descripción verbal, obtención de información cualitativa e interpretación de elementos significativos de gráficos sencillos relativos a fenómenos cercanos.

– *Segundo ciclo:* Iniciación al uso de estrategias eficaces de recuento de datos; recogida y registro de datos sobre objetos, fenómenos y situaciones familiares utilizando técnicas elementales de encuesta, observación y medición. Disposición a la elaboración y presentación de gráficos y tablas de forma ordenada y clara.

– *Tercer ciclo*: Distintas formas de representar la información. Tipos de gráficos estadísticos. Valoración de la importancia de analizar críticamente las informaciones que se presentan a través de gráficos estadísticos. La media aritmética, la moda y el rango, aplicación a situaciones familiares. Obtención y utilización de información para la realización de gráficos.

Para estudiar el problema de investigación planteado, en este trabajo analizaremos los gráficos elaborados por una muestra de 93 futuros profesores de educación primaria al trabajar con un proyecto abierto de análisis de datos. Los gráficos producidos se clasifican, en primer lugar, atendiendo a su complejidad semiótica, que se define a partir del análisis de los objetos matemáticos utilizados en su elaboración, relacionándola con la corrección de los gráficos. Se analiza, asimismo, la capacidad de los profesores en la interpretación de los gráficos construidos por ellos mismos y en la extracción de conclusiones sobre el problema de investigación planteado en el proyecto. En lo que sigue, primeramente se describe los antecedentes de la investigación, seguidamente la metodología del estudio, para finalizar con la presentación y discusión de los resultados, así como de sus implicaciones en la formación de profesores.

ANTECEDENTES

A pesar de la importancia señalada de la competencia relacionada con las gráficas estadísticas, la investigación didáctica nos alerta que dicha competencia no se alcanza, en general, en la educación obligatoria (Cazorla, 2002), produciéndose errores en las escalas (Li y Shen, 1992) o en la construcción de gráficos específicos (Pereira Mendoza y Mellor, 1990; Lee y Meletiou, 2003; Bakker, Biehler y Konold, 2004).

El gráfico, como objeto semiótico

Una posible explicación de este hecho es que la simplicidad del lenguaje gráfico es aparente, pues incluso el más elemental de los gráficos puede considerarse, de acuerdo a diversos autores, como un objeto semiótico complejo. Por ejemplo, Bertin (1967) asume la premisa de que un gráfico es un texto multimodal; tanto en su conjunto como en los elementos que lo componen están constituidos por conjuntos de signos que requieren una actividad semiótica por aquellos que los interpretan. Para este autor, la lectura de un gráfico comienza con una *identificación externa* del tema al que se refiere, a través de la interpretación del significado del título y las etiquetas. A continuación se requiere una *identificación interna* de las dimensiones relevantes de variación en el gráfico, es decir, la interpretación de las variables representadas y sus escalas. Finalmente, se produce una *percepción de la correspondencia* entre los niveles particulares de cada dimensión visual, para obtener conclusiones sobre los niveles particulares de cada variable y sus relaciones en la realidad representada.

Observamos que, en cada uno de los pasos descritos por Bertin en la lectura de un gráfico, se puede identificar una o varias funciones semióticas en el sentido de Eco (1977), quien las define como correspondencia entre un antecedente (expresión) y un consecuente (contenido), establecida por un sujeto. En la lectura de gráficos el estudiante debe realizar varias actividades de traducción entre el gráfico en su conjunto o una parte del gráfico y lo representado. La interpretación correcta requiere conocimientos no siempre disponibles por el estudiante sobre los convenios de construcción y elementos del gráfico, que son los siguientes (Curcio, 1987; 1989):

– Las *palabras* que aparecen en el gráfico, como su título, las etiquetas de los ejes y de las escalas, que proporcionan las claves necesarias para comprender las relaciones representadas.

– El *contenido matemático* subyacente, por ejemplo los conjuntos numéricos empleados, el área en un diagrama de sectores, longitud en un gráfico de líneas, coordenadas en un diagrama de dispersión, que el estudiante ha de dominar para interpretarlo.

– Los *convenios específicos* que se usan en cada gráfico y que se deben conocer para poder realizar una lectura o construcción correcta. Por ejemplo, el alumno ha de conocer en un diagrama de sectores que la amplitud del sector es proporcional a la frecuencia; en un diagrama de dispersión, que cada punto representa un caso y las coordenadas del punto los valores de las dos variables representadas.

En relación con los anteriores componentes Friel, Curcio y Bright (2001) describen las siguientes competencias relacionadas con el lenguaje gráfico:

– *Reconocer* los elementos estructurales del gráfico (ejes, escalas, etiquetas, elementos específicos) y sus relaciones. Distinguir si cada elemento es o no apropiado en el gráfico particular.

– *Apreciar el impacto* de cada uno de estos componentes sobre la presentación de la información (por ejemplo, predecir cómo cambiaría el gráfico al variar la escala de un eje).

– *Traducir las relaciones* reflejadas en el gráfico a los datos que se representan en el mismo y viceversa.

– *Reconocer cuándo un gráfico es más útil que otro* en función del juicio requerido y de los datos representados, es decir, saber elegir el gráfico adecuado al tipo de variable y al tipo de problema.

Niveles de comprensión

Además de las competencias anteriores, algunos autores definen niveles en la lectura crítica de datos y muestran que no todos los alumnos alcanzan el nivel más alto. Bertin (1967) definió los siguientes:

– *Extracción de datos*, que consiste en poner en relación un elemento de un eje con el de otro eje. Por ejemplo, en un diagrama de barras leer la frecuencia asociada a un valor de la variable.

– *Extracción de tendencias,* cuando se es capaz de percibir en el gráfico una relación entre dos subconjuntos de datos que pueden ser definidos a priori o visualmente. Un caso particular es determinar visualmente la moda de una distribución en un diagrama de barras, ya que se clasifican los datos en subconjuntos (que tienen un mismo valor para la variable) y se comparan entre sí estos subconjuntos para ver cuál tiene mayor frecuencia.

– *Análisis de la estructura* de los datos, comparando tendencias o agrupamientos y efectuando predicciones. Un ejemplo ocurre cuando se representa en un diagrama de barras adosadas dos distribuciones y se analizan sus diferencias en promedios y dispersión.

Si un gráfico permite uno de estos niveles de lectura, también posibilita los inferiores a él, pero no al contrario. Bertin define una *imagen* como una forma visual que permite percibir en un solo instante una correspondencia representada en el gráfico (por ejemplo, la barra de un diagrama de barras es una imagen, pues permite apreciar la frecuencia que corresponde a un valor de la variable). Un gráfico sería para el autor tanto más eficaz cuanto mayor es el nivel de lectura que permite con una sola imagen.

Otra clasificación de niveles de lectura de gráficos muy similar a la de Bertin y que ha tenido un gran impacto en educación estadística se debe a Curcio (1989), quien mostró que las principales dificultades aparecen en los niveles superiores y que el nivel de lectura progresa con la edad de los estudiantes. Denomina a los tres niveles definidos por Bertin *«leer los datos»* (lectura literal del gráfico sin interpretar la información contenida en el mismo), *«leer dentro de los datos»* (interpretación e integración de los datos en el gráfico) y «*leer más allá de los datos*» (predicciones e inferencias a partir de los datos sobre informaciones que no se reflejan directamente en el gráfico). Friel, Curcio y Bright (2001) amplían la clasificación definiendo un nuevo nivel *«leer detrás de los datos»* consistente en valorar críticamente el método de recogida de datos, su validez y fiabilidad, así como las posibilidades de extensión de las conclusiones.

Más recientemente, estos niveles se extendieron para tener en cuenta la valoración crítica de la información, una vez alcanzada la lectura completa del gráfico (Aoyama y Stephen, 2003; Aoyama, 2007):

– *Nivel racional/literal*. Los estudiantes leen correctamente el gráfico, interpolan, detectan tendencias y predicen. Para responder la pregunta planteada, usan las características del gráfico, pero no cuestionan la información ni dan explicaciones alternativas.

– *Nivel crítico*. Los estudiantes leen los gráficos, comprenden el contexto y evalúan la fiabilidad de la información, cuestionándola, pero son incapaces de buscar hipótesis que expliquen la disparidad entre el gráfico y la conclusión.

– *Nivel hipotético*. Los estudiantes leen los gráficos, los interpretan y evalúan la información, formando sus propias hipótesis y modelos alternativos a los sugeridos en la pregunta planteada.

Competencias gráficas de los futuros profesores

Bruno y Espinel (2005) analizan la forma en que futuros profesores construyen un histograma de frecuencias a partir de una lista de datos e indican que las dificultades con los gráficos estadísticos también se presentan en los futuros profesores. Aproximadamente la mitad de los participantes en su estudio tuvieron errores, incluyendo la representación de los intervalos de variación de la variable en el eje de ordenadas, la omisión de intervalos de frecuencia nula, o el uso de rectángulos no adosados en variables continuas. En cuanto al polígono de frecuencias, los futuros profesores no unen las marcas de clase, omiten el intervalo de frecuencia nula o confunden la frecuencia y el valor de la variable.

Continuando la investigación anterior, Bruno y Espinel compararon los errores de los futuros profesores en la construcción del histograma y el polígono de frecuencias con la evaluación de los producidos por posibles estudiantes. Prácticamente todos los futuros profesores cometieron algún error al construir los gráficos, pero lo más preocupante fue la falta de coherencia entre su construcción del gráfico y la forma en que evaluaron las respuestas de estudiantes ficticios. Además, en caso de coherencia, generalmente se trataba de futuros profesores que cometieron errores en la interpretación de los gráficos y también consideraron correctos los gráficos incorrectos de sus posibles estudiantes.

Preocupadas por estos resultados, las autoras continúan la investigación utilizando un cuestionario que trata de evaluar la cultura y razonamiento estadístico de los futuros profesores por medio de su interpretación de gráficos, comparando los resultados con los de otros estudiantes universitarios americanos (Espinel, 2007). Aunque en ambos grupos de estudiantes las tareas fueron difíciles, la dificultad fue mayor para los futuros profesores españoles, sobre todo al predecir la forma de un gráfico a partir de la descripción verbal de variables conocidas por los estudiantes o al leer los histogramas.

Monteiro y Ainley (2006; 2007) indican que la lectura de gráficos en el contexto escolar es una tarea más limitada que la posible interpretación de dichos gráficos en otras actividades de la vida diaria. La razón dada por los autores es que, mientras en la escuela sólo pedimos a los estudiantes una respuesta correcta desde el punto de vista matemático, en contextos extraescolares intervienen también otros conocimientos no matemáticos. Monteiro y Ainley estudiaron la competencia de futuros profesores en la lectura de gráficos tomados de la prensa diaria, encontrando que muchos no tenían conocimientos matemáticos suficientes para llevar a cabo dicha lectura. La mayoría de los profesores participantes no tuvieron formación específica en la lectura de gráficos estadísticos y reconocieron sus carencias al respecto. En esta investigación también se observó que la interpretación de los gráficos moviliza conocimientos y sentimientos que inciden en su comprensión. Por ejemplo, se obtuvo mucho mejores resultados al interpretar un gráfico sobre incidencia de cáncer en las mujeres que otro matemáticamente equivalente en un contexto de menor interés para el alumno.

Observamos que las investigaciones centradas en la competencia gráfica de los profesores en su mayoría se interesan únicamente por la competencia y niveles alcanzados en la lectura de los gráficos y no por la competencia o niveles alcanzados en su construcción. Aunque en sus trabajos Espinel y sus colaboradores estudian los gráficos construidos por futuros profesores, se centran en el análisis de los errores cometidos o la consistencia entre el gráfico que construyen y la forma en que evalúan un gráfico como correcto o incorrecto.

Además, las tareas usadas por estos autores son cerradas, indicándoseles a los futuros profesores el gráfico que han de construir y no hay preguntas relacionadas con la interpretación posterior del gráfico. En nuestro trabajo nos interesamos por el nivel de complejidad de los gráficos producidos por los futuros profesores en un proyecto abierto de análisis de datos. También nos interesamos por la interpretación que los futuros profesores hacen de los gráficos construidos y por las conclusiones que obtienen sobre el problema de investigación planteado en el proyecto.

METODOLOGÍA

Muestra

Los estudiantes participantes en el estudio (n = 93, divididos en tres grupos) cursaban el segundo o tercer año del plan de estudio de la Diplomatura de Magisterio, en la especialidad de Educación Primaria. Habían cursado una asignatura de «Matemáticas y su Didáctica» en el primer curso de la Diplomatura de Magisterio y en ella habían estudiado los gráficos, tablas y resúmenes estadísticos básicos (10 horas lectivas), trabajando también con un proyecto sencillo de análisis de datos, aunque en aquel caso los datos fueron dados por el profesor, en lugar de ser recogidos por los alumnos.

La mayoría de los estudiantes de la muestra procedían de bachillerato de letras, en el que los contenidos matemáticos son menores que en otras especialidades, aunque los que provienen de bachillerato de ciencias sociales habrían estudiado estadística por un cuatrimestre completo, el curso anterior a su ingreso en la Facultad de Educación. Durante la Enseñanza Primaria y Secundaria Obligatoria todos los estudiantes habrían estudiado los gráficos estadísticos elementales.

Contexto y tarea propuesta

Los datos se tomaron dentro de una de las prácticas de un curso de Currículo Matemático en la Facultad de Educación, Universidad de Granada, en la que se propuso la realización de un proyecto de análisis de datos por parte de los futuros profesores.

La pregunta de investigación planteada en el proyecto (tomado de Batanero (2001) fue evaluar las intuiciones del conjunto de estudiantes de cada grupo sobre los experimentos aleatorios. Para ello, se realizó en la clase y en cada uno de los grupos uno de los experimentos utilizados en la investigación sobre percepción de la aleatoriedad (ver, por ejemplo, Serrano, 1996 o Nickerson, 2002). En el experimento se pidió a cada estudiante inventar una secuencia de 20 lanzamientos de una moneda equilibrada (sin lanzarla realmente). Los futuros profesores realizaron individualmente el experimento, inventando una secuencia de 20 lanzamientos (secuencia simulada) y anotaron los resultados en una hoja de registro, escribiendo C para cara y + para cruz. A continuación, cada estudiante obtuvo otra secuencia de 20 lanzamientos de una moneda, lanzando esta vez realmente la moneda y anotando asimismo los resultados en la hoja de registro (secuencia real). Se pidió a los estudiantes que contasen el número total de caras en cada una de las dos secuencias, y al finalizar la clase el profesor proporcionó a los estudiantes una hoja de datos que contenía para cada alumno el número de caras de las secuencias real y simulada (Ver datos obtenidos en uno de los grupos en la tabla 1).

La tarea propuesta a los futuros profesores fue analizar individualmente los datos recogidos en clase en el experimento (una tabla de datos similar a la tabla 1) y producir un informe escrito en el que debían comparar las distribuciones del número de caras en las secuencias real y simulada y justificar, en base al análisis de los datos, si la clase en su conjunto tenía o no buenas intuiciones sobre el azar. Los estudiantes tuvieron libertad para elegir los gráficos o resúmenes estadísticos que considerasen convenientes e incluso para usar ordenadores. A la semana siguiente se recogieron y analizaron las producciones de los estudiantes.

Más detalles de la actividad se recogen en Godino, Batanero, Roa y Wilhelmi (2008).

Análisis a priori de la tarea

Observamos que en el proyecto planteado a los futuros profesores aparecen dos variables aleatorias y sus correspondientes variables estadísticas:

– La variable η_r: número de caras en 20 lanzamientos de la moneda equilibrada, que es una variable aleatoria binomial con parámetros $n = 20$ $p = q = 1/2$. Su media es igual a np y su varianza igual a npq.

– La variable aleatoria η_s: número de caras en una secuencia de longitud 20 inventada por los estudiantes y que usamos como modelo matemático para reflejar las intuiciones colectivas sobre los experimentos aleatorios. No sería una variable binomial, pues la investigación didáctica (por ejemplo, Serrano, 1996) muestra que los ensayos producidos por las personas en este tipo de experimentos no son independientes, aunque la probabilidad de cara sigue siendo $p = 1/2$.

Además, puesto que cada estudiante repite el experimento y cuenta el número de caras obtenidas, tenemos una serie de m ensayos de cada una de las variables aleatorias anteriores, siendo m el número de estudiantes de la muestra, y por lo tanto obtenemos dos variables estadísticas:

– La variable estadística Y_r o resultados de una muestra de m valores de la variable aleatoria η_r. El estudiante ha de usar esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria η_r puesto que sus conocimientos matemáticos no le permiten trabajar directamente con η_r.

– La variable estadística Y_s o resultados de una muestra de m valores de la variable aleatoria η_s. El estudiante ha de usar esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria η_s puesto que sus conocimientos matemáticos no le permiten trabajar directamente con η_s.

– Más concretamente, el estudiante ha de comparar las características de las distribuciones de las variables Y_r e Y_s para conjeturar sobre las semejanzas o diferencias entre η_r y η_s y de acá deducir sobre las intuiciones respecto al azar.

En la tabla 2 presentamos los estadísticos del número de caras en las dos distribuciones de las variables estadísticas Y_r *e* Y_s y en la figura 1 su representación conjunta. Se espera que los estudiantes elaboren un gráfico similar al de la figura 1 u otro equivalente (por ejemplo, un gráfico conjunto de líneas o de puntos). También se espera que, al interpretar la gráfica, observen que los valores de las medidas de tendencia central (la moda o mediana se aprecian visualmente) son muy similares y en consecuencia los alumnos han tenido buena intuición respecto al promedio del número de caras al lanzar 20 veces la moneda.

Tabla 1

Número de caras obtenidas por cada alumno en las secuencias reales y simuladas al lanzar 20 monedas.

ALUMNO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
SIMULADA	7	10	11	10	11	10	11	13	8	11	11	10	12	10	11	10	12	9	12	10	10	7	9	10	11	10	12	10	11	13	11	12
REAL	7	15	14	9	13	8	11	9	12	12	12	10	11	11	11	12	11	7	14	8	10	6	9	13	9	8	11	11	11	10	10	10

Por otro lado, la mayor variabilidad del número de caras en las secuencias reales (que se observa del análisis del rango en las dos variables) indica que la intuición del conjunto de estudiantes en la clase respecto a la dispersión es pobre y se tiende a producir secuencias aleatorias con menor variabilidad a la esperada teóricamente (conclusiones similares se han obtenido en las investigaciones sobre percepción subjetiva de la aleatoridad reseñadas en Serrano, 1996 o Nickerson, 2002). Alternativamente o adicionalmente a la producción de gráficos, los estudiantes podrían resolver el problema calculando y comparando alguna de las medidas de tendencia central y dispersión presentadas en la tabla 2 para obtener las mismas conclusiones.

Figura 1
Distribución del número de caras en las secuencias real y simulada.

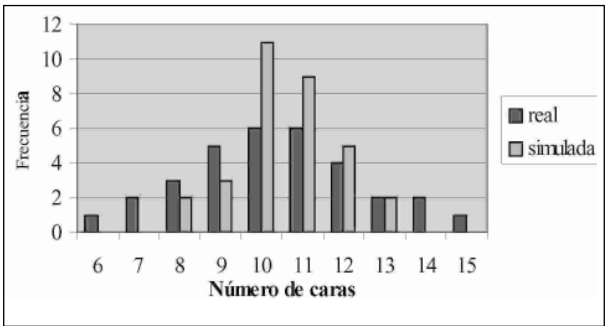


Tabla 2
Estadísticos del número de caras en las secuencias real y simulada.

	SECUENCIA SIMULADA	SECUENCIA REAL
Media	10,4	10,4
Mediana	10,5	10
Moda	10	10,11
Varianza	2	4,3
D. típica	1,4	2,1
Mínimo	8	6
Máximo	13	15
Rango	5	9

RESULTADOS

Una vez recogido el informe del análisis de datos del proyecto que entregó cada estudiante, se analizaron los gráficos producidos por los estudiantes desde varios puntos de vista.

Complejidad semiótica de los gráficos producidos

En primer lugar, se realizó un análisis de las variables representadas en el gráfico y otros objetos matemáticos que nos permitió definir un «*nivel de complejidad semiótica*» de elaboración propia. Se definió este nivel de complejidad teniendo en cuenta los objetos matemáticos cuyo conocimiento requiere el sujeto para construir el gráfico, así como el nivel de lectura (en la clasificación de Bertin, 1967) que posibilita el gráfico construido.

Siguiendo a Font, Godino y D'Amore (2007), en nuestro trabajo asumimos que en las prácticas matemáticas que realizan los sujetos al resolver problemas se presentan múltiples funciones semióticas (bien de lectura o de representación), debido a la necesidad de usar y operar con objetos matemáticos, que son inmateriales. Estos autores consideran una tipología de objetos matemáticos (expresiones verbales o simbólicas, propiedades, procedimientos, problemas, argumentos, conceptos) que intervienen en las prácticas matemáticas y cada una de las cuales puede jugar el papel de antecedente o consecuente de una función semiótica.

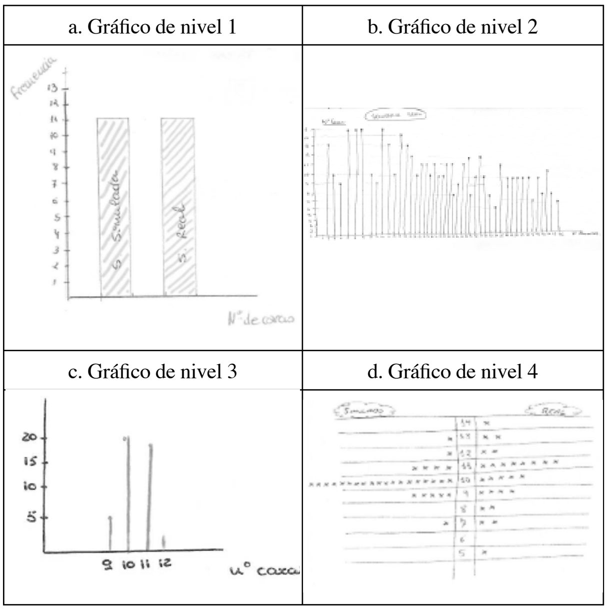
La tarea propuesta a los futuros profesores en nuestra investigación (comparar las dos distribuciones dadas para decidir sobre las intuiciones del conjunto de alumnos) constituye para ellos un problema, pues la solución no les es inmediata. Las prácticas realizadas por los futuros profesores para resolverlo y, específicamente, el gráfico producido ponen en juego una serie de objetos matemáticos que varían en los diferentes gráficos. Por tanto, varían las funciones semióticas subyacentes a la construcción e interpretación de estos gráficos por parte de los estudiantes, así como en la puesta en relación de la pregunta planteada en el proyecto con el gráfico construido, por medio de una argumentación.

La actividad semiótica realizada por los futuros profesores con relación a estos gráficos comienza durante su construcción, puesto que deben representar en el gráfico tanto objetos matemáticos (inmateriales) como los resultados del experimento realizado en la clase (que pertenecen al plano de la realidad). A partir de la lista dada de datos (Tabla 1), el sujeto que construye el gráfico, primero ha de considerar los datos individuales como valores de una variable (establece una correspondencia entre cada dato de la tabla con el objetos matemático «valor» y de cada fila de la tabla con el objeto matemático «variable»). A su vez, cada dato de la tabla remite a un resultado de un experimento hecho en la clase (elemento fenomenológico). El alumno ha de fijar uno de los ejes del gráfico para representar esta variable y sus posibles valores (debe establecer una correspondencia entre los valores numéricos de la variable y la posición de puntos en el eje).

Por otro lado, un mismo tipo de gráfico (por ejemplo, un gráfico simple de barras) se puede usar para representar diferentes objetos matemáticos, tales como frecuencias absolutas, relativas, porcentajes y frecuencias acumuladas, medias u otros resúmenes estadísticos. La regla de correspondencia que explicita cuál es el objeto matemático representado en el gráfico deberá ser precisada en las etiquetas y escalas de los ejes. El título del gráfico proporcionará la clave para interpretar la realidad modelizada por los objetos matemáticos representados en el gráfico.

Dependiendo del gráfico construido se usarán diferentes objetos matemáticos, y la actividad semiótica implicada será más o menos compleja. Por ello, los gráficos producidos no deben considerarse simplemente como representaciones equivalentes de un concepto subyacente (la distribución de datos obtenida) sino como configuraciones diferenciadas de objetos relacionados e interactuando con dicha distribución, en términos de Godino, Font y D'Amore (2007). Estas consideraciones nos sirvieron para definir niveles de complejidad semiótica en los gráficos producidos, que se describen a continuación. En la figura 1 presentamos un ejemplo de gráficos correctos producidos en cada una de los niveles 1 a 4 (en el caso de los niveles 2 y 3 el alumno elabora otro gráfico para el número de caras de la secuencia real). Algunos futuros profesores no produjeron gráficos, se limitaron a realizar cálculos estadísticos, en la mayoría de los casos reducidos a las medidas de posición central (media, mediana y/o moda); a veces también la dispersión (rango, desviación típica).

Figura 2
Ejemplos de gráficos producidos en cada nivel de complejidad.



Nivel 1. Representa sólo sus resultados individuales. Algunos alumnos producen una gráfica donde únicamente representan los datos obtenidos en su experimento particular, sin considerar los de sus compañeros. Estos estudiantes tratan de resolver la pregunta para su caso particular (si él mismo tiene una buena intuición), no habiendo comprendido el propósito del proyecto o no siendo capaces de realizar un análisis global de los datos.

La actividad semiótica realizada es la más elemental posible, dentro del proyecto planteado, pues sólo representan un dato (una columna de la tabla 1, en lugar de representar la tabla completa). Un ejemplo se presenta en la figura 2a. El estudiante representa en un diagrama de barras los valores obtenidos en el lanzamiento de la moneda en sus 20 ensayos en las secuencias real y simulada. Aunque emplea explícitamente la palabra «frecuencia», no está representando una variable, sino los resultados (constantes) que obtuvo en sus secuencias. Implícitamente aplica el concepto «frecuencia» a la variable Bernoulli «resultado de cada experimento individual» que toma los valores: «Cara» o «cruz». Pero no considera la variable «número de caras en la secuencia de cada alumno de la clase» que tomaría valores numéricos. El alumno está representando en el eje Y los números naturales; establece, por tanto, una función semiótica entre los puntos del eje y los numerales en él representado y también entre estos puntos y numerales y los números naturales (hasta 13).

La gráfica producida sólo permite un nivel de lectura de *extracción de datos* (Bertin, 1967) o *leer los datos* en la categorización de Curcio (1989), pues sólo permite responder preguntas sobre la frecuencia de caras en cada una de las dos secuencias del alumno.

Nivel 2. Representa los valores individuales de la variable. Algunos estudiantes representan todos los datos de la tabla 1, en el orden en que aparecen en la tabla, sin llegar a agrupar los valores iguales de la variable «número de caras al lanzar 20 monedas» en las secuencias reales o simuladas. Un ejemplo se presenta en la figura 2b. Al representar todos los valores obtenidos, el alumno está implícitamente usando la idea de variable. De hecho hace una referencia explícita a la misma en el rótulo «n.º caras» del eje *Y* (expresión verbal) que remite a un concepto (variable) y a la particularización de dicho concepto en el proyecto (el número de caras es variable). El alumno utiliza el eje *Y* para representar los valores de la variable. Asimismo, representa sobre el eje los números naturales (hasta 15); establece, por tanto, una función semiótica entre los puntos del eje y los numerales en él representado y también entre estos numerales y los números naturales (hasta 15).

Por otro lado, el alumno utiliza el eje *X* para representar los diferentes alumnos de la clase, asignando a cada uno de ellos un número natural consecutivo situado sobre dicho eje. Cada uno de los puntos marcados y numerados sobre este eje (por ejemplo el «2») remite a uno de los estudiantes (el segundo estudiante en la lista de datos) y también al valor de la variable obtenido por dicho estudiante (que se representa con una barra situada sobre el punto asignado al estudiante y cuya altura es el valor del número de caras obtenidas por él). Hacemos notar que el orden de presentación de los datos en el eje *X* es artificial, pues sólo indica el orden arbitrario en que se recogieron los datos en la clase. Tampoco se calcula ni usa la idea de frecuencia asociada a cada valor de la variable o de distribución de frecuencias de la variable.

El nivel de lectura de los gráficos que hemos clasificado en el nivel 2 es superior al de los de nivel 1, pues permite visualizar todos los valores obtenidos de la variable, así como su variabilidad (se observa claramente el valor máximo y mínimo). Sin embargo, no se llega a percibir claramente la estructura o tendencia de los datos. Por tanto, la gráfica permitiría plantear preguntas de nivel de «extracción de datos» pero no permite todavía la «extracción de tendencias» (Bertin, 1967), pues no es posible percibir en el gráfico una relación entre dos o más subconjuntos de datos.

Nivel 3. Produce gráficos separados para cada distribución. El alumno representa un gráfico para cada una de las dos variables «número de caras en las secuencias reales» y «número de caras en las secuencias simuladas», incluyendo cada uno de los valores diferentes de la variable con sus frecuencias. En este caso (ver ejemplo en figura 2c) el alumno agrupa todos los casos con el mismo valor de cada una de las variables y representa cada uno de los valores diferentes de la variable una única vez en el eje *X*. De este modo los puntos y numerales representados en dicho eje remiten a los valores que ha tomado la variable en el experimento y a su rango de variación. El orden del eje *X* es ahora el orden numérico natural; los valores de la variable se representan siguiendo este orden (y no el orden de aparición en los experimentos efectuados en la clase).

El estudiante asigna un número (frecuencia) para «representar» un subconjunto de datos (el conjunto de casos que tienen el mismo valor de la variable; y el cardinal de dicho conjunto). Se usa el eje *Y* para representar las frecuencias (aunque en el ejemplo no se explicita cuál es el objeto matemático representado en este eje). La representación gráfica del conjunto de valores de la variable así establecida con sus frecuencias remite a un nuevo objeto matemático, la *distribución de frecuencias,* concepto complejo, que se refiere al agregado de todos los datos obtenidos por el conjunto de alumnos de la clase y no a ninguno de los datos de un alumno en particular. Es decir, en los niveles 3 y 4 se representa una función que a cada valor de la variable asigna su frecuencia (nivel 3) o una frecuencia en cada variable (nivel 4). Esta función no llega a construirse en los niveles anteriores.

Los gráficos que hemos clasificado en el nivel 3, permiten responder preguntas hasta el nivel de *extracción de tendencias* en la categorización de Bertin (1967), pues en este tipo de gráficos es posible percibir directamente una relación entre dos o más subconjuntos de datos. Comparando las frecuencias de las diferentes barras del ejemplo (Figura 3c) se puede determinar el de mayor frecuencia y percibir directamente la moda, además de ser también patente la variabilidad, que ahora se visualiza más claramente aún que en el gráfico de nivel 2, examinando directamente el rango de variación de la variable en el eje *X*.

Nivel 4. Produce un gráfico conjunto de las dos distribuciones. Aunque en los gráficos de nivel 3 ya está presente la distribución de la variable, al usar dos gráficos separados se dificulta su comparación, sobre todo cuando el alumno usa diferentes escalas en cada gráfico. Un nivel más alto de complejidad se alcanza cuando el alumno ha llegado a formar las distribuciones de las dos variables y las representa conjuntamente en el mismo gráfico, estrategia que facilitará la comparación (ver ejemplo en figura 2d). Aunque las frecuencias absolutas no aparecen explícitamente representadas en este ejemplo, implícitamente también se consideran, por el número de puntos. Los ejes *X* e *Y* estarían intercambiados respecto al gráfico de barras. Además de toda la actividad semiótica descrita en la construcción del gráfico de nivel 3, el estudiante ha de seleccionar un rango de variación y escala común para las dos distribuciones y establecer algún convenio en el gráfico que permita diferenciar una variable de otra (en el ejemplo, la colocación de los puntos a izquierda o derecha de un eje central; también la dirección en que se van añadiendo puntos al gráfico).

Los gráficos que hemos clasificado en el nivel 4 permiten el nivel superior de lectura en la categorización de Bertin (1967), es decir, el *análisis de la estructura* pues permite comparar tanto tendencias como variabilidad en las dos variables en una única imagen. En resumen, aunque dentro de cada uno de los niveles que hemos definido en los gráficos producidos por los estudiantes observamos una variedad de gráficos, se evidencia un salto cualitativo entre cada uno de los niveles definidos. Por un lado, cada nivel en la construcción de gráficos implica la actividad semiótica usada en el nivel anterior y además la amplía, como también amplía la complejidad y el número de objetos matemáticos utilizados en la comprensión.

En términos de Font, Godino y D'Amore (2007), la complejidad de la configuración de objetos matemáticos subyacente crece con el nivel del gráfico producido. Por otro lado, se puede establecer un paralelismo entre los niveles de lectura de gráficos descritos por Bertin (1967) y asumidos posteriormente por Curcio (1989) con otra terminología y nuestros niveles de construcción de los gráficos: El nivel 1 posibilita la extracción de datos; el nivel 3, la extracción de tendencias y el 4 el análisis de la estructura; mientras que el 2 permite un nivel intermedio superior a la simple extracción de datos pero sin llegar a la extracción de tendencias.

Relación entre complejidad y corrección del gráfico

En la tabla 3 presentamos la distribución de alumnos en función del nivel de gráfico elaborado, su corrección o no y la interpretación obtenida. Del total de 93 alumnos 88 (94,6%) producen algún tipo de gráfico para analizar sus datos, incluso cuando las instrucciones de la tarea no los requerían. Este alto porcentaje indica la necesidad sentida de los estudiantes de producir un gráfico y llegar, mediante un proceso de transnumeración (Wild y Pfannkuch, 1999), a un conocimiento no disponible en los datos brutos.

Tabla 3
Clasificación de estudiantes según nivel de complejidad semiótica y corrección de los gráficos.

NIVEL DE COMPLEJIDAD SEMIÓTICA	CORRECCIÓN DEL GRÁFICO			TOTAL EN EL NIVEL
	Corr.	Parcial	Inc.	
N1. Representa sólo sus datos	1		1	2
N2. Representa resultados individuales	10	1	4	15
N3. Gráficos separados para cada muestra	15	17	14	46
N4. Gráficos conjuntos	14	6	5	25
Total	40	24	24	88

La mayoría de los que elaboran gráficos (52,2%), producen gráficos de nivel 3, es decir, representan la distribución de cada una de las dos variables por separado (gráficos de barras horizontales y verticales y polígonos de frecuencias). Generalmente son correctos o parcialmente correctos (formalmente correctos pero usan diferente escala en los dos gráficos o bien diferente gráfico para cada variable; no centran el intervalo en los histogramas, no hay coincidencia de los valores representados con la escala utilizada o no incluyen etiquetas en los ejes). Catorce alumnos en este nivel elaboran gráficos incorrectos al representar en un mismo gráfico variables no comparables, representar los productos de valores por frecuencia o bien intercambian frecuencias y valores de las variables en los ejes, considerando las frecuencias como variable independiente y los valores de la variable como dependiente, error ya detectado en Ruiz (2006) en relación con el concepto de variable aleatoria. Los errores de los gráficos reproducen los descritos por Espinel (2007) en su investigación.

El 28,4% de los estudiantes trabajan al nivel 4 y producen un solo gráfico de las dos variables, aunque seis de estos gráficos son parcialmente correctos y cinco incorrectos por alguna de las razones señaladas anteriormente.

Son pocos los estudiantes que analizan sólo sus propios datos (nivel 1) y sólo el 17% estudian los valores obtenidos por cada estudiante caso a caso sin llegar a formar la distribución (nivel 2), por lo que el concepto de distribución parece ser adquirido y utilizado por los estudiantes para resolver la tarea propuesta. Dentro del nivel 2 hemos obtenido diagramas de barras horizontales y verticales; gráficos de líneas de una o las dos variables que, aunque no permiten resolver el problema de comparación, al menos muestran la variabilidad de los datos en los diferentes estudiantes. Otros estudiantes producen gráficos claramente inapropiados, que ni siquiera permiten visualizar la variabilidad de los datos, entre ellos, diagramas de sectores, gráficos adosados o apilados de barras.

No se observa, en consecuencia, una relación clara entre el nivel alcanzado en la construcción del gráfico y su corrección, porque, aunque la mayoría de gráficos producido en el nivel 4 son correctos, también lo son la mayoría de los gráficos producidos en el nivel 2. Sin embargo, estos últimos no permiten resolver el problema propuesto.

Nivel de lectura de los gráficos

En la tabla 4 clasificamos a los estudiantes según nivel del gráfico y lectura que hacen del mismo. Destacamos, en primer lugar, que una cuarta parte de los futuros profesores se limitan a producir el gráfico y lo presentan sin ningún comentario sobre el mismo o bien hacen errores en su lectura. Otros estudiantes, aunque leen elementos aislados del gráfico (dominan la extracción de datos), no llegan al nivel de extracción de tendencias o análisis de la estructura (Bertin, 1967) en su lectura de los gráficos construidos.

Los errores de lectura se producen, sobre todo, en los gráficos de nivel 3 y 4, pues la menor complejidad semiótica de los gráficos de nivel 1 y 2 hace que las preguntas que los estudiantes puedan plantearse sobre los mismos sean sólo de nivel de extracción de datos (Bertin, 1967), nivel que es más asequible.

Tabla 4
Clasificación de estudiantes según nivel de complejidad semiótica y lectura de los gráficos.

NIVEL DE COMPLEJIDAD SEMIÓTICA	INTERPRETACIÓN DEL GRÁFICO			TOTAL EN EL NIVEL
	CORR.	PARCIAL	BLANCO O INC.	
N1. Representa sólo sus datos	1	1		2
N2. Representa resultados individuales	4	10	1	15
N3. Gráficos separados para cada muestra	15	15	16	46
N4. Gráficos conjuntos	9	11	5	25
Total	29	37	22	88

Un 42% de los estudiantes hacen una interpretación parcialmente correcta de los gráficos, analizando tan sólo los promedios sin tener en cuenta la variabilidad o bien al contrario, comparando sólo los rangos de variación, sin tener en cuenta las tendencias. La mayor proporción relativa de los que hacen esta interpretación parcial se encuentra entre los estudiantes que producen un gráfico de nivel 1 que, como hemos indicado, no llegan a trabajar con la distribución de la variable estadística.

La explicación es que la comprensión de la idea de distribución entraña conjugar a la vez las ideas de promedio y dispersión (Bakker y Gravemeijer, 2004), que los estudiantes en este nivel no han sabido todavía relacionar. Una tercera parte de los estudiantes conjuga estas dos ideas detectando las tendencias y analizando la estructura de los datos, aumentando la proporción de los que lo hacen en los dos niveles superiores del gráfico construido.

Conclusiones sobre el problema planteado

Completada la extracción de tendencias y análisis de la estructura del gráfico, habría que relacionarlas con la pregunta planteada en el proyecto, esto es, ver qué indican sobre las intuiciones de los estudiantes en relación con los fenómenos aleatorios. Se trataría de hacer una valoración crítica de la información del gráfico, una vez finalizada su lectura (Aoyama y Stephen, 2003; Aoyama, 2007).

En la tabla 5 relacionamos el nivel de complejidad semiótica del gráfico o gráficos producido y las conclusiones obtenidas en relación con las intuiciones de la clase respecto a los fenómenos aleatorios. Observamos que la obtención de la conclusión es la tarea más difícil para todos los estudiantes, siendo sólo una tercera parte los que obtienen una conclusión, al menos parcial, y creciendo esta proporción con el nivel de complejidad semiótica del gráfico. Ello es lógico, pues los niveles descritos por Aoyama y Stephen comienzan a partir del nivel superior de Bertin (1967) y Curcio (1989). Es preciso, primero, obtener las tendencias y estructura de los datos para poder valorarlos críticamente.

Todos los estudiantes con gráficos en nivel 1 y la mayoría de los que están a nivel 2 hacen una conclusión incorrecta o bien no concluyen. La mayor parte de las conclusiones incorrectas son debidas a una intuición errónea sobre la aleatoriedad, suponiendo que una buena intuición implicaría que la secuencia simulada por un estudiante fuese idéntica en alguna característica a su secuencia real. Estos estudiantes no perciben la impredecibilidad de los experimentos aleatorios y tratan de predecir sus resultados o al menos predecir una parte de los resultados de la secuencia aleatoria en el experimento realizado en clase. Este fenómeno, denominado *ilusión de control,* fue observado originalmente de Langer (1975) en diferentes tipos de juegos de azar, por ejemplo la lotería, donde observó en los sujetos una creencia ilusoria sobre su capacidad para influir en el resultado.

Tabla 5
Clasificación de estudiantes según nivel de complejidad semiótica y conclusiones sobre el problema planteado.

NIVEL DE COMPLEJIDAD SEMIÓTICA	CONCLUSIÓN			TOTAL EN EL NIVEL
	CORR.	PARCIAL	BLANCO O INC.	
N1. Representa sólo sus datos			2	2
N2. Representa resultados individuales		3	12	15
N3. Gráficos separados para cada muestra	1	12	33	46
N4. Gráficos conjuntos	1	7	17	25
Total	2	22	64	88

Cincuenta estudiantes con gráficos de nivel 3 y 4 han hecho errores en la lectura de datos de extracción de tendencias o análisis de estructura, por lo que llegan a una conclusión incorrecta o bien, habiendo hecho una lectura correcta, se encuentran en el *nivel racional/literal* en la clasificación de Aoyama y Stephen (2003). Aunque leen correctamente el gráfico, interpolan, detectan tendencias y predicen, son incapaces de buscar hipótesis que expliquen la disparidad o semejanza entre los dos gráficos y no llegan a una conclusión.

Veintidós estudiantes llegan a una conclusión parcial, habiendo alcanzado parcialmente el nivel hipotético descrito por Aoyama y Stephen (2003). Estos estudiantes indican que las intuiciones son buenas, pues el promedio del número de caras se aproxima al valor esperado 10, sin tener en cuenta los resultados obtenidos al comparar las medidas de dispersión. Estos estudiantes muestran una buena comprensión de la idea de esperanza matemática (media de la variable aleatoria) y la discriminan de la media de la correspondiente variable estadística.

Sólo dos estudiantes concluyen sobre la diferencia en dispersión en las dos distribuciones, es decir, que el grupo tiene buena intuición respecto al promedio de número de caras, aunque, al ser las secuencias reales más variables que las simuladas, las intuiciones sobre la variabilidad de los fenómenos aleatorios es pobre en los estudiantes. Además de mostrar una comprensión completa de la idea de distribución y un alto nivel en la lectura y construcción de gráficos, usarían de una forma consistente el nivel hipotético de valoración crítica de la información en la clasificación de Aoyama y Stephen (2003).

DISCUSIÓN Y CONCLUSIONES

El nivel de complejidad semiótica que hemos propuesto en este trabajo ha permitido clasificar los gráficos producidos por los futuros profesores en una tarea abierta. Complementa, por tanto, el estudio de la actividad semiótica relacionada con los gráficos estadísticos y las categorías de niveles de lectura propuestos por autores como Bertin (1967), Curcio (1989) y Aoyama y Stephen (2007). Es también consistente con estas categorizaciones, en cuanto un nivel mayor de complejidad semiótica en un gráfico en nuestra clasificación posibilita el planteamiento de un nivel superior de lectura en las mismas. Por otro lado, al asumir en nuestro trabajo la multiplicidad de objetos matemáticos que, de acuerdo a Font, Godino y D'Amore (2007) pueden jugar el papel de expresión o contenido en una función semiótica, nuestro análisis pone de manifiesto que la complejidad semiótica del gráfico se define no sólo por el nivel de lectura que permite, sino sobre todo por el número y complejidad de objetos matemáticos implícitos en el mismo.

La investigación reseñada muestra también que la construcción e interpretación de gráficos es una habilidad altamente compleja, y confirma las dificultades descritas por Bruno y Espinel (2005) y Espinel (2007) en futuros profesores, a pesar de que han de transmitir el lenguaje gráfico a sus alumnos y utilizarlo como herramienta en su vida profesional. También amplía el trabajo de las autoras proporcionando datos sobre la capacidad de construcción de gráficos de los futuros profesores en una tarea abierta y mostrando que la mayoría de los participantes no consigue elaborar un gráfico de complejidad suficiente para permitir resolver el problema. El concepto de distribución, esencia del razonamiento estadístico según Wild y Pfannkuch (1999), no llega a ser utilizado por una parte de los futuros profesores y el razonamiento sobre la variabilidad, que es otro de los componentes esenciales del razonamiento estadístico, según estos autores es difícil para la mayoría.

En el proyecto planteado los estudiantes recorren todos los pasos del método estadístico, desde el planteamiento del problema, la definición de las preguntas, recogida y análisis de datos y obtención de conclusiones. También se pone en práctica el proceso de modelización, pues según Henry (1997) «*un modelo es una interpretación abstracta, simplificada e idealizada de un objeto del mundo real, de un sistema de relaciones o de un proceso evolutivo que surge de una descripción de la realidad*» (p. 78). En nuestro proyecto la realidad se ha simplificado y abstraído pasando de la idea general de intuición al experimento concreto y la definición de variables aleatorias (número de caras en las secuencias real y simuladas). Además de trabajar con las variable aleatorias y estadísticas correspondientes, los estudiantes han de interpretar los resultados del trabajo matemático realizado con el modelo (distribuciones de datos obtenidas) en el contexto del problema (traducir estos resultados a lo que indican respecto de las intuiciones de los estudiantes).

Es precisamente este último paso (puesta en relación del resultado con la pregunta planteada) el que ha causado más dificultad, al no alcanzar los niveles crítico e hipotético (Aoyama y Stephen, 2003) en la interpretación crítica de la información. La dificultad también se explica por la falta de familiaridad de los futuros profesores con proyectos estadísticos y actividades de modelización. Puesto que estas actividades se recomiendan hoy en la enseñanza de la estadística en Educación Primaria y pueden ser especialmente adecuadas en el trabajo individual y en grupos recomendados en el marco del Espacio Europeo de Educación Superior, pensamos que deberían emplearse en la formación de profesores.

En consecuencia, sería necesario atender a estos problemas en la formación de los profesores de Educación Primaria, pues una mejora de la educación de los niños pasa por la formación del profesor, que no debe olvidar el lenguaje de las gráficas estadísticas.

AGRADECIMIENTOS

Este trabajo es parte del Proyecto SEJ2007-0110/EDUC, MEC-FEDER y beca FPU AP2007-03222

REFERENCIAS BIBLIOGRÁFICAS

AOYAMA, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education, 2*(3). On line <www.iejme/>.

AOYAMA, K.M. y STEPHENS, M. (2003). Graph interpretation aspects of statistical literacy: A Japanese perspective. *Mathematics Education Research Journal, 15*(3), 3-22.

BAKKER, A., BIEHLER, R. y KONOLD, C. (2004). Should young students learn about box plots?, en Burrill, G. y Camden, M. (eds.). *Curricular Development in Statistics Education. Proceedings of the: International Association for Statistical Education (IASE) Roundtable,* pp. 163-173. Auckland: IASE. On Line <www.stat.auckland.ac.nz/~iase/publications/>.

BAKKER, A. y GRAVEMEIJER, K.P.E. (2004). Learning to reason about distribution, en Garfield, J. y Ben Zvi, D. (eds.). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking,* pp 147-168. Dordrecht, The Netherlands: Kluwer.

BATANERO, C. (2001). *Didáctica de la estadística*. Granada. Grupo de Educación Estadística.

BERTIN, J. (1967). *Semiologie graphique*. París: Gauthier-Villars.

BRUNO, A. y ESPINEL, M.C. (2005). Recta numérica, escalas y gráficas estadísticas: un estudio con estudiantes para profesores. *Formación del Profesorado e Investigación en Educación Matemáticas, 7,* pp. 57-85.

CAZORLA, I. (2002). *A relaçao entre a habilidades viso-pictóricas e o dominio de conceitos estatísticos na leitura de gráficos*. Tesis Doctoral. Universidad de Campinas.

CURCIO, F.R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education,* 18(5), pp. 382-393.

CURCIO, F.R. (1989). *Developing graph comprehension*. Reston, VA: N.C.T.M.

ECO, U. (1977). *Tratado de semiótica general*. Barcelona: Lumen.

ESPINEL, C. (2007). Construcción y razonamiento de gráficos estadísticos en la formación de profesores. *Investigación en Educación Matemática,* 11, pp. 99-119.

FONT, J.D., GODINO, J.D. y D'AMORE, B. (2007). An ontosemiotic approach to representations in mathematics education. *For the Learning of Mathematics*, 27(2), pp. 3-9.

FRIEL, S., CURCIO, F. y BRIGHT, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education*, 32(2), pp. 124-158.

GAL, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review* 70(1), pp. 1-25.

GODINO, J.D., BATANERO, C., ROA, R. y WILHELMI, M.R. (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work, en Batanero, C., Burrill, G., Reading, C. y Rossman,A. (eds.). *Proceedings of the Joint ICMI / IASE StudyTeaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education*. Monterrey, México: ICMI e IASE. On line: <www.stat.auckland.ac.nz/~iase/publicatons>.

HENRY, M. (1997). Notion de modéle et modélization en l'enseignement, en *Enseigner les probabilités au lycée,* pp. 77-84. Reims: Commission Inter-IREM-

LANGER, E.J. (1975). The illusion of control. *Journal of Personality and Social Psychology,* 32, pp. 311-328.

LEE, C. y MELETIOU, M. (2003). Some difficulties of learning histograms in introductory statistics. *Joint Statistical Meetings - Section on Statistical Education*. On line <www.statlit.org/PDF/2003LeeASA.pdf>.

LI, D.Y. y SHEN, S.M. (1992). Students'weaknesses in statistical projects. *Teaching Statistics,* 14(1), pp. 2-8.

MEC (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Primaria*.

MONTEIRO, C. y AINLEY, J. (2006). Student teachers interpreting media graphs, en Rossman, A. y Chance, B. (eds.). *Proceedings of the Seventh International Conference on Teaching Statistics*. Salvador de Bahia: International Statistical Institute and International Association for Statistical Education. Online <www.stat.auckland.ac.nz/~iase>.

MONTEIRO, C. y AINLEY, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education,* 2(3), pp. 188-207. On line <www.iejme/>.

NICKERSON, R.S. (2002). The production and perception of randomness. *Psychological Review,* 109(2), pp. 330-357.

PEREIRA-MENDOZA, L. y MELLOR, J. (1990). Student's concepts of bar graph: Some preliminary findings. *Proceedings of the Third International Conference on Teaching Statistics* Ed D. Vere-Jones. Voorburg: International Statistical Institute.

RUIZ, B. (2006). *Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria*. Tesis de Maestria. CICATA. México.

SCHIELD, M. (2006). Statistical literay survey analysis: reading graphs and tables of rates percentages, en Phillips, B. (ed.). *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town: International Statistical Institute and International Association for Statistical Education. On line <www. stat.auckland.ac.nz/~iase>.

SERRANO, L. (1996). *Significados institucionales y personales de conceptos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad*. Tesis doctoral. Universidad de Granada.

WILD, C. y PFANNKUCH, M. (1999). Statistical thinking in empirical enquiry (con discusión). *International Statistical Review,* 67(3), pp. 223-265.

[Artículo recibido en mayo de 2008 y aceptado en junio de 2009]

Analysing the Semiotic Complexity of Graphs Produced by Future Primary School Teachers in a Task Related to the Comparison of Two Statistical Variables

Batanero, Carmen[1]; Arteaga, Pedro[1] y Ruiz, Blanca[2]
[1] Departamento de Didáctica de las Matemáticas. Universidad d Granada
[2] Departamento de Didáctica de las Matemática. Instituto tecnológico de Monterrey
batanero@ugr.es
pedroarteagacezon@yahoo.com
bruiz@itesm.mx

Summary

Graphical language is essential in analysing data, a tool for transnumeration, and a component of statistical literacy. Because recent curricular guidelines in Spain introduce statistics graphs in primary school level, our research was aimed at assessing the graphical competence of prospective primary school teachers. Previous research suggests that these prospective teachers make errors when building graphs and lack coherence in evaluating graphs carried out by fictitious students.

To complement this research, we analysed the graphs produced by 93 prospective primary school teachers when working in an open statistical project where they had to compare two statistical variables coming from binomial distributions. The data were collected along a mathematics education course (second year of University) in the Faculty of Education, University of Granada and participants had studied descriptive statistics (graphs, tables, averages, spread) the previous academic year.

We carried out a qualitative analysis of the reports provided by the participants and classified the graphs produced taking into account the type of graph, number of variables represented in the graph, and the mathematical objects involved in its construction. Using ideas from the onto-semiotic approach to research developed by Godino, Batanero, Contreras, Font and Wilhelmi in different papers, we also analysed the semiotic activity involved in building and interpreting the graphs, and defined different levels of semiotic complexity as follows:

L1. Representing only the student's individual results. Some students only displayed in the graph the data they obtained in his/her particular experiment, without considering all the data in the classroom. These students showed an intuitive idea of statistical variable and distribution since they represented the frequency of results in their individual experiment. However they only considered the Bernoulli variable the "result of an individual experiment", instead of considering a Binomial distribution "number of successes in *n* experiments".

L2. Representing the individual values for each student. These students used all the data but did not group the similar results in the different experiments carried out by the different students. Instead, they represented the value (or values) obtained by each student in the classroom in the order the data were collected, so they didn't compute the frequency of the different values or explicitly use the idea of distribution. The order of data in the X-axis was artificial since it only indicated the arbitrary order in which the students were located in the classroom.

L3. Producing separate graphs for each distribution. The student produced a separate graph for each of the two variables with each of the different values of the variable and its frequency. These students went from the data set to the statistical variable and its distribution and used the ideas of frequencies and distribution. The order in the X-axis was the natural order in the real line.

L4. Producing a joint graph for the two distributions. The students formed the distributions for the two variables and represented them in a joint graph, which facilitated the comparison; the graph was more complex than those in level L3, since it represented two different variables.

Out of a total of 93 students, 88 produced some graphs when analysing the data, even if the instructions given to the student did not explicitly require that they construct a graph. This fact suggests that students felt the need to build a graph and reached, by a transnumeration process some information that was not available in the raw data. 81% of the students that produced a graph, produced level 3 or 4 graphs, that were generally correct o partly correct (correct graph with different scales or different graph in each sample; not centring the rectangles in the histogram, or missing labels). Consequently the concept of distribution seemed natural for the majority of students who used it to solve the task, although the instructions did not require this explicitly.

In general, these prospective teachers interpreted correctly or partially correctly the graphs in all the levels, reaching Curcio's (1989) reading between the data level. However, only 1/3 of students who built level 3 and 4 graphs reached the «reading between the data» level, the remaining students only made a partial interpretation or did not interpret the graph at all. Only a few students got a final conclusion, in relating the result with the research question.

We conclude that building and interpreting graphs is a complex activity for prospective teachers, even though they should transmit graphical language to their students. Improving the teaching of statistics in schools should begin with teacher education that should take into account statistical graphs.