ANÁLISIS DE GRÁFICOS ESTADÍSTICOS ELABORADOS EN UN PROYECTO DE ANÁLISIS DE DATOS

Pedro Arteaga

Tesis de Máster

Universidad de Granada, 2009

Directora: Carmen Batanero

Depósito Legal: GR 763-2009

ISBN: 978-84-691-7511-8

Publica:

Grupo de Investigación en Educación Estadística

Departamento de Didáctica de la Matemática

Universidad de Granada

Imprime:

Servicio de Reprografía de la Facultad de Ciencias

Universidad de Granada

Avda. Fuentenueva s/n 18071 Granada

Financiación:

Trabajo realizado en el marco de la beca FPU AP2007-03222 y Proyecto SEJ2007-60110 /EDUC, MEC-Feder.

ÍNDICE

1. INTRODUCCIÓN

Asistimos en la actualidad a un incremento de los contenidos de estadística que se recomiendan en la escuela primaria, hecho que se hace patente en los Decretos de Enseñanzas Mínimas de la Educación Primaria (MEC, 2006a). En dichos Decretos se dispone que la Educación Primaria tiene carácter obligatorio y gratuito y comprende seis cursos organizados en tres ciclos de dos cursos cada uno, debiéndose incorporar los alumnos al primer curso el año natural en el que cumplan seis años. Una de las áreas de conocimiento obligatorias que se impartirán en los tres ciclos es la de Matemáticas, y se observa que desde primer ciclo se incluye dentro de dicha área un nuevo Bloque (Bloque 4) llamado *Tratamiento de la información, azar y probabilidad*, indicándose en este que:

> ñLos contenidos adquieren su pleno significado cuando se presentan en conexión con actividades que implican a otras áreas de conocimiento. Igualmente el trabajo ha de incidir de forma significativa en la comprensión de las informaciones de los medios de comunicación, para suscitar el interés por los temas y ayudar a valorar el beneficio que los conocimientos estadísticos proporcionan ante la toma de decisiones, normalmente sobre cuestiones que estudian otras áreas. Tienen especial importancia en el bloque los contenidos actitudinales, que favorecen la presentación de los datos de forma ordenada y gráfica, y permiten descubrir que las matemáticas facilitan la resolución de problemas de la vida diaria. A su vez, los contenidos de este bloque deben iniciar en el uso crítico de la información recibida por diferentes mediosò (pg. 43096).

Observamos que, además de los contenidos tradicionales, y de iniciar antes la enseñanza de los temas estadísticos, hay también un cambio en el enfoque, recomendándose el desarrollo del razonamiento estadístico de los niños y la presentación de la estadística como un instrumento para resolver problemas y no sólo como un conjunto de técnicas.

Uno de los objetivos concretos es utilizar las técnicas elementales de recogida de datos para obtener información sobre fenómenos y situaciones del entorno; representarla de forma gráfica y numérica y formarse un juicio sobre la misma. En relación con los gráficos, el trabajo con estos se incluye dentro de los contenidos del Bloque

Tratamiento de la información, azar y probabilidad en todos los ciclos de la Educación primaria. En el primer ciclo se comienza con interpretaciones de determinados elementos de un gráfico sencillo relacionado con fenómenos cercanos a los niños y progresivamente se pasa a los contenidos del tercer ciclo en los que se estudiarán distintos tipos de gráficos estadísticos y se deberá conseguir que los niños aprecien la importancia que tiene el poder valorar críticamente informaciones que son presentadas a través de gráficos. Además se manifiesta que la destreza en la utilización de representaciones gráficas para interpretar la información aporta una herramienta muy valiosa para conocer y analizar mejor la realidad. Como criterio de evaluación para el primer ciclo se indica:

> ñRealizar interpretaciones elementales de los datos presentados en gráficas de barras. Formular y resolver sencillos problemas en los que intervenga la lectura de gráficos. Con este criterio se trata de valorar la capacidad de interpretar gráficos sencillos de situaciones familiares y verificar la habilidad para reconocer gráficamente informaciones cuantificablesò (pg. 43098)

Para que estas propuestas puedan llevarse a cabo, será necesario preparar a los futuros maestros tanto en lo que respecta al conocimiento estadístico, como en el conocimiento pedagógico relacionado con su enseñanza. Esta es una preocupación asumida por organismos internacionales como la IASE (International Association for Statistical Education) e ICMI (International Commission on Mathematical Instruction) que han lanzado un estudio internacional enfocado sobre la problemática de formación del profesorado en estadística, puesto que el problema no es exclusivo de España (http://www.ugr.es/~icmi/iase_study/).

Estos son algunos de los motivos que nos han llevado a interesarnos en la competencia gráfica estadística de los futuros maestros de Educación Primaria, así, en este trabajo presentamos una investigación inicial relacionada con este tema, en concreto se analizan las gráficas estadísticas construidas por una muestra de futuros profesores en el desarrollo de un proyecto abierto de análisis de datos. Comenzamos contextualizando el problema, mediante la descripción de sus objetivos, el contexto educativo y la justificación del uso de proyectos en la enseñanza de la estadística.

A continuación presentamos un estado de la cuestión sobre las investigaciones relacionadas con la comprensión de los gráficos estadísticos, que será necesario completar en el futuro, pero nos permite introducirnos en el tema.

En el estudio empírico que sigue se hace un análisis a priori del proyecto propuesto, y se analizan las gráficas producidas por los futuros profesores desde diversos puntos de vista. Finalizamos con la presentación de nuestras conclusiones y algunas publicaciones que han resultado de esta investigación.

2. CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

2.1. INTRODUCCIÓN

Para contextualizar nuestro trabajo comenzaremos haciendo un breve resumen de algunos elementos teóricos que se van a utilizar y que forman parte del Enfoque Ontosemiótico desarrollado por Godino y sus colaboradores.

Seguidamente especificamos nuestros objetivos e hipótesis y resaltamos el interés, tanto para la formación de profesores, como para aportar nuevo conocimiento en el área de didáctica de la estadística.

El resto del capítulo describe el interés del trabajo con proyectos en la clase de estadística y su contribución al desarrollo de algunas competencias que se incluyen en los nuevos Decretos de Enseñanzas Mínimas. La recogida de los datos en nuestra investigación se hace a partir de un informe realizado por los estudiantes en un proyecto de análisis de datos. Por otro lado, el uso de proyectos se recomienda hoy día en varias investigaciones e incluso se deja entender en los mismos Decretos.

2.2. ELEMENTOS TEÓRICOS

En este trabajo vamos a utilizar algunas nociones teóricas desarrolladas por Godino y colaboradores en el marco teórico llamado Enfoque Ontosemiótico (EOS), que nos van a resultar de utilidad para analizar las producciones gráficas realizadas por una muestra de futuros profesores en un proyecto abierto de análisis de datos, que describiremos detalladamente en el punto 4. Además, estos elementos teóricos van a permitirnos definir una jerarquía en la construcción de los gráficos estadísticos en función de la complejidad semiótica de estos.

En diversas publicaciones, los autores que han desarrollado este marco teórico se interesan por los objetos matemáticos que surgen de las prácticas matemáticas realizadas para resolver un problema que puede ser matemático o extra-matemático (Godino, 2002; Godino, Batanero y Font, 2007). En nuestro trabajo, plantearemos a los futuros profesores un problema extra-matemático cuya finalidad es decidir si las intuiciones del conjunto de alumnos de la clase respecto a la aleatoriedad son correctas o hasta qué punto lo son. Este problema extra-matemático llevará a organizar en la clase un experimento que cada alumno realizará individualmente y a modelizar la situación, para así definir algunas variables estadísticas y recoger datos sobre las mismas.

Así, del problema extra-matemático se pasará a un problema matemático consistente en la comparación de dos variables estadísticas, para decidir sobre la igualdad o desigualdad de sus distribuciones. Nuestro interés se centra en las prácticas matemáticas de los estudiantes para resolver, primero el problema matemático, y, seguidamente el extra-matemático, y en concreto en su elaboración e interpretación de gráficos.

Godino, Batanero y Font (2007) indican que de los sistemas de prácticas matemáticas operativas y discursivas emergen nuevos objetos, que serán objetos institucionales si los sistemas de prácticas son compartidos en el seno de una institución o serán objetos personales si dichos sistemas de prácticas son realizados por una persona. Es decir, se tienen en cuenta las dimensiones social y personal del conocimiento y también el hecho de que una práctica personal pudiera ser o no adecuada desde el punto de vista de la institución.

Para permitir analizar con detalle los procesos didácticos, en el EOS se formula una ontología de los objetos matemáticos que tiene en cuenta el triple aspecto de la matemática como actividad socialmente compartida de resolución de problemas, como lenguaje simbólico y como sistema conceptual lógicamente organizado. A continuación se muestran las categorías de objetos o entidades matemáticas primarias propuestas desde el EOS. Cualquiera de estas entidades puede ser objeto de análisis didáctico. Así en nuestro caso nos centramos en los gráficos estadísticos, mientras otras investigaciones podrían centrarse en los problemas, algún concepto o proposición, etc.

- *Situaciones-problemas*: aplicaciones extra-matemáticas, ejercicios, problemas, acciones que inducen una actividad matemática. En nuestro caso la situación planteada es el estudio de las intuiciones de los estudiantes (situación extra-matemática) que se transforma durante su resolución en un problema matemático (comparación de dos series de datos). Para resolver este problema matemático y el extra-matemático relacionado los estudiantes harán una serie de prácticas, como efectuar un recuento de los datos y construir gráficas estadísticas. En el análisis que haremos en el punto 4 no consideramos las situaciones-problemas, ya que son fijas para todos los estudiantes y para todos los gráficos analizados.

- *Lenguajes*: términos, expresiones, notaciones, gráficos que se utilizan para representar los datos del problema, las operaciones que hacemos con ellos, los objetos matemáticos que se utilizan y la solución encontrada. Nosotros nos interesaremos por los gráficos producidos por los estudiantes que por un lado, serían en sí mismos, parte del lenguaje matemático. Pero, por otro, se pueden considerar a su vez como objetos matemáticos complejos, dentro de los cuales aparecen diversos tipos de lenguaje. Esperamos que en los gráficos construidos los alumnos mezclen el lenguaje verbal y gráfico, aunque también pudieran usar el simbólico. Este tipo de lenguajes se organizan dentro del gráfico en función de varios elementos del mismo: su título, ejes, escalas, marcas del eje y elementos geométricos específicos del gráfico.
- *Conceptos- definición*: En las prácticas que llevan a cabo los estudiantes para resolver un problema matemático (en este caso cuando resuelven el problema planteado o construyen los gráficos) se usan implícita o explícitamente objetos matemáticos, de los cuáles el alumno ha de recordar o aplicar la definición. Por ejemplo, los estudiantes usarán implícitamente los objetos: números enteros y decimales, fracciones, variable estadística, valor, rango de la variable, frecuencia y en algunos tipos de gráficos la proporcionalidad, sistema de coordenadas cartesianas, orden numérico, longitud, segmento u otros.
- *Proposiciones* o enunciados sobre relaciones o propiedades de los conceptos que igualmente se han de emplear al resolver problemas matemáticos. Por ejemplo, cuando los estudiantes han de usar la relación de proporcionalidad entre una frecuencia y la altura de una barra cuando construyen el diagrama de barras o para comprobar si su representación es completa han de recordar el hecho de que la suma de frecuencias ha de ser igual al tamaño de la muestra o incluso cuando detectan visualmente la moda, han de buscar el valor (o valores) de la variable con mayor frecuencia.
- *Procedimientos*: Serían los algoritmos, operaciones, técnicas de cálculo que los estudiantes han aprendido durante la enseñanza previa y que aplican al resolver el problema. En nuestro ejemplo, los alumnos aplican algoritmos de representación de números enteros en la recta numérica, ordenación y recuento de datos. En el punto 4 haremos un análisis detallado del lenguaje, conceptos, proposiciones y procedimientos empleados en algunos de los gráficos producidos por los estudiantes.

- *Argumentos*: Serían los enunciados usados para validar o explicar las proposiciones y procedimientos o bien la solución de los problemas. Los estudiantes en nuestro estudio usan argumentos para comparar los datos o justificar si las intuiciones de los estudiantes son o no correctas. Nosotros analizaremos separadamente los argumentos de los estudiantes en los apartados 4.4.3 y 4.4.4.

Otro punto de interés en nuestro estudio es el de función semiótica que sirve para resaltar los procesos de interpretación que se llevan a cabo en la actividad matemática y en los cuales a veces pueden aparecer desajustes (conflictos) de interpretación entre alumnos y profesor. Partiendo de diversos autores, Font, Godino y DôAmore (2007) describen la idea de función semiótica como correspondencia entre un antecedente y un consecuente, establecida por un sujeto (persona o institución). Un gráfico estadístico, en sí mismo, puede considerarse como una función semiótica, donde el antecedente es el propio gráfico y lo representado es la distribución estadística de los datos, siendo la correspondencia el conjunto de convenios establecidos en estadística para el gráfico particular, que permite a la persona que lee el gráfico interpretarlo o bien a la persona que tiene los datos construirlo. En este ejemplo se pone de manifiesto la posibilidad de que una persona no conozca los convenios citados y por tanto no interprete o interprete incorrectamente el gráfico, produciéndose un conflicto que los autores denominan semiótico.

El antecedente (expresión) y consecuente (significado) de una función semiótica no se restringen a conceptos, sino abarca toda la anterior ontología de objetos matemáticos u organizaciones de estos en entidades más complejas como sistemas conceptuales o teorías. Así en nuestro estudio, podemos descomponer el gráfico (que como hemos dicho es un objeto matemático complejo) y centrarnos en partes del mismo, observando una multiplicidad de funciones semióticas dentro del gráfico. Por ejemplo, si el título de un gráfico es ñDistribución del número de caras en la secuencia realò observamos que algunas palabras de este título (lenguaje) hacen referencia a una serie de conceptos matemáticos (distribución, número, secuencia).

Dichas funciones semióticas juegan un papel fundamental en los procesos matemáticos que se realizan en cualquier práctica matemática, en los que se relacionan entidades primarias mostradas anteriormente, o grupo de ellas. Algunos de los procesos matemáticos citados por los autores son los de comunicación, problematización, definición, enunciación, elaboración de procedimientos (algoritmización, rutinización,.) y argumentación.

Las entidades u objetos primarios pueden relacionarse entre sí formando lo que dichos autores denominan ñconfiguracionesò, definidas como redes de los objetos intervinientes y emergentes en las prácticas matemáticas y las relaciones entre ellos. Hay dos tipos de configuraciones: configuración epistémica (redes de objetos institucionales) y configuración cognitiva (redes de objetos personales).

En nuestro trabajo al analizar las producciones gráficas estadísticas realizadas por los alumnos lo que analizaremos son las configuraciones cognitivas, que es una herramienta potente que consiste en un desglose de las prácticas de los alumnos en entidades situacionales, lingüísticas, actuativas, conceptuales, proposicionales y argumentativas que permiten un análisis más fino del aprendizaje matemático llevado a cabo por estos. En nuestro trabajo esto nos permitirá identificar la estructura de los objetos que han hecho posible la realización del gráfico por parte del alumno (práctica matemática), identificar incluso conflictos semióticos y llegar a definir una clasificación de los gráficos elaborados por los alumnos según unos niveles de complejidad semiótica que definiremos en el apartado 4.

2.3. OBJETIVOS DE LA INVESTIGACIÓN

Como se ha señalado, nuestro trabajo se interesa por la formación estadística de los futuros profesores de primaria y más concretamente por su capacidad en la construcción de gráficos estadísticos en una tarea abierta.

La importancia de la formación en estadística y su didáctica de los futuros profesores de primaria se deduce claramente del papel asignado a la Estadística en los Decretos de Enseñanzas Mínimas para la Educación Primaria y de los errores detectados en conocimientos estadísticos elementales en futuros profesores que se han descrito, por ejemplo en Estepa (1993), Bruno y Espinel (2005) o Espinel (2007). Como ya hemos indicado estos problemas se producen en otros países y han llevado a la organización de un estudio internacional (Joint ICMI/IASE Study).

Dentro de este objetivo general, los objetivos específicos que nos proponemos son los siguientes:

Objetivo 1. Iniciar un estado de la cuestión de las investigaciones sobre comprensión de gráficas estadísticas, especialmente las que se relacionan con la formación de profesores.

Este objetivo se hace necesario para poder iniciar una investigación, no sólo para el trabajo de Master, sino cara a una futura tesis doctoral. Los trabajos que hemos encontrado sobre la comprensión de gráficos estadísticos son dispersos y no conocemos ninguna investigación tipo survey sobre este punto.

Objetivo 2. Identificar un problema de investigación de interés dentro de la formación de profesores y en el campo de la estadística

Puesto que tenemos intención de continuar la investigación para tratar de hacer una tesis doctoral, será importante encontrar un problema no investigado, que sea de interés para la educación estadística y tenga suficiente entidad para permitirnos realizar una tesis en el futuro.

Objetivo 3. Explorar algún instrumento de obtención de datos de futuros profesores sobre su capacidad de construcción de gráficos.

Tratamos de adquirir experiencia de investigación y esto incluye también realizar una primera toma de datos. Hemos elegido, entre otros posibles instrumentos un proyecto abierto de análisis de datos.

La exploración también permitirá tener algún criterio de evaluación de los proyectos abiertos realizados por los futuros profesores, de modo que pueda utilizarse para evaluar la enseñanza de la estadística a través de proyectos, además de cómo útil de investigación.

Objetivo 4. Recoger algunos gráficos elaborados por futuros profesores con dicho instrumento y clasificarlos para definir una jerarquía inicial de niveles de complejidad en la construcción de dichos gráficos.

Con ello adquiriremos competencia en el análisis de datos cualitativos y produciremos una primera categorización de los gráficos que en el futuro se pueda mejorar y pueda ser útil tanto para la enseñanza como para la investigación. Este es el objetivo de mayor interés en el estudio, en cuanto no hemos encontrado investigaciones que describan niveles en la construcción de gráficos y de este modo aportaríamos un resultado original.

2.4. HIPÓTESIS DE LA INVESTIGACIÓN

Una primera revisión literaria sobre la competencia gráfica de futuros profesores de educación primaria, nos muestra que los futuros profesores tienen dificultades en la construcción de gráficos estadísticos (Bruno y Espinel, 2005) e incluso en la lectura e interpretación de estos (Monteiro y Ainley, 2006,2007). En nuestra investigación pretendemos clasificar los gráficos construidos por futuros profesores de Educación Primaria en un proyecto de análisis de datos según distintos niveles de complejidad semiótica, y teniendo en cuenta las dificultades mostradas en distintas investigaciones por los futuros profesores en este tipo de tareas, nuestras hipótesis de investigación son las siguientes:

Hipótesis 1. Los futuros profesores de nuestra muestra, al construir los gráficos estadísticos para llevar a cabo la tarea propuesta, cometen muchos de los errores ya detectados en investigaciones sobre dificultades y errores en la construcción de los gráficos estadísticos.

En nuestra investigación los alumnos participantes en el estudio seleccionarán ellos mismos los gráficos que crean pertinentes para resolver el problema que se les planteó, por ello esperamos encontrar una variedad de distintos tipos de gráficos y muchos de los errores ya encontrados en distintos estudios sobre la construcción de los distintos tipos de gráficos estadísticos (Espinel, 2007; Espinel Bruno y Plasencia, 2008).

Hipótesis 2. Los dos niveles superiores de complejidad semiótica definidos en nuestra investigación no son alcanzados por la mayoría de los estudiantes de la muestra de nuestro estudio.

Al observar que la competencia en la lectura de los gráficos estadísticos es más difícil de alcanzar de lo que podría pensarse, según se indica en algunas investigaciones previas (Friel, Curcio y Bright, 2001; Aoyama, 2007), pensamos que los estudiantes de nuestra muestra no llegarán a los niveles más altos de complejidad en la construcción de sus gráficos.

2.5. CONTEXTO EDUCATIVO

Nuestra investigación se llevó a cabo como una práctica de la asignatura ñCurrículo de Matemáticas en la Educación Primariaò, en la Diplomatura de Magisterio. Esta es una asignatura de carácter eminentemente práctico (dos tercios de las horas docentes se dedican a actividades prácticas) cuyo contenido exclusivo es la Didáctica de la Matemática, entendida respecto a la matemática incluida en el currículo de Educación Primaria en España, que abarca cuatro bloques temáticos: *Números y operaciones, La medida: estimación y cálculo de magnitudes, Geometría, Tratamiento de la información, azar y probabilidad* (MEC, 2006a).

La actividad de la cual obtuvimos nuestros datos formaba parte de una práctica a la que se dedicó dos sesiones de clase de dos horas de duración cada una y se relaciona con el cuarto bloque del currículo de matemáticas en la educación primaria: *Tratamiento de la información, azar y probabilidad* teniendo, además, los estudiantes que realizar un desarrollo personal de la actividad y producir un informe escrito individual fuera del horario lectivo. Este requisito es parte de la metodología de enseñanza universitaria implantada en la Facultad de Ciencias de la Educación para adecuar los planes de estudio al Espacio Europeo de Educación Superior (EEES), iniciativa seguida por los países europeos para homogeneizar los títulos profesionales y hacerlos válidos en todo el territorio europeo. Ello implica similitud de contenidos y metodología de enseñanza, disminuyendo la parte dedicada a lecciones magistrales y aumentando el trabajo individual y en grupos de los estudiantes.

Los estudiantes en cuestión habían cursado también una asignatura de ñMatemáticas y su Didácticaò en el primer curso de la Diplomatura de Magisterio (el año anterior a la experiencia), asignatura de contenido exclusivamente matemático, aunque restringida a las matemáticas que el futuro maestro tendrá que enseñar en la Educación Primaria. También, como su nombre indica, se trata de relacionar en todo momento la matemática, con la enseñanza de la misma, aunque la parte puramente didáctica se deja para la asignatura de ñCurrículo Matemático en la Educación Primariaò a que hemos hecho referencia.

Dentro de los contenidos de ñMatemáticas y su Didácticaò (asignatura de 9 créditos, cada uno de los cuales equivale a 25 horas de trabajo del estudiante, incluido el trabajo dentro y fuera del aula) se dedicaron 2 créditos (50 horas del estudiante) al bloque de *Tratamiento de la información, azar y probabilidad.* Los estudiantes revisaron y ampliaron su información sobre datos, variables estadísticas, distribuciones de frecuencias, gráficos sencillos (diagramas de barras, polígonos de frecuencias, histogramas, gráficos de sectores), medidas de posición central (media, mediana y moda) y medidas de dispersión (rango y desviación típica). También se revisaron y ampliaron las ideas de aleatoriedad, probabilidad y la asignación de probabilidades mediante regla de Laplace o probabilidad frecuencial. No se hizo un estudio formal de la variable aleatoria.

Los estudiantes habían trabajado también en ñMatemáticas y su Didácticaò con un proyecto sencillo de análisis de datos (aunque una sola vez y los datos no fueron recogidos personalmente, sino que fueron dados por el profesor). Hay que tener también en cuenta que los estudiantes proceden de diferentes bachilleratos, siendo en su mayoría alumnos provenientes de bachillerato de humanidades y ciencias sociales, en los cuales los contenidos matemáticos son menores que en los otros bachilleratos, aunque hay mayor contenido de estadística.

En todo caso, cualquiera que sea el bachillerato cursado, durante la Enseñanza Primaria (6-12 años) y Enseñanza Secundaria Obligatoria (12-16) todos los estudiantes habrán recibido una enseñanza mínima de estadística que incluye los mismos conceptos trabajados el primer año de la Facultad en la asignatura de Matemáticas y su Didáctica.

2.6. TRABAJO CON PROYECTOS EN LA CLASE DE ESTADÍSTICA

2.6.1. INTRODUCCIÓN

En algunos trabajos previos (Díaz, Arteaga y Batanero, 2007; Díaz y Arteaga, 2008) hemos analizado las posibilidades que brinda el trabajo con proyectos en la enseñanza de la estadística para contribuir a la adquisición de las competencias básicas incluidas en los recientes decretos de enseñanza mínimas para la Educación Secundaria Obligatoria (MEC, 2006b). Creemos importante incluir acá alguna de estas ideas puesto que la recogida de datos se hace a través de un proyecto.

El interés del trabajo con proyectos en la clase de estadística ha sido sugerido anteriormente por muchos otros autores. Por ejemplo Biehler (1997) indica que ayudan a superar la distancia entre la comprensión de los conceptos y los medios técnicos de cálculo para poder aplicarlos.

Batanero y Díaz (2004) sugieren que la mejor forma de lograr la cultura estadística de los estudiantes es introducir en las clases de estadística el trabajo con proyectos, pues, en lugar de introducir los conceptos y técnicas descontextualizadas, o aplicadas únicamente a problemas como los que aparecen en los libros de texto, consistentes en calcular un promedio o hacer un gráfico y que son difíciles de encontrar en la vida real, un proyecto permite presentar las diferentes fases de una investigación estadística: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado.

Los proyectos aumentan la motivación de los estudiantes (Holmes, 1997), puesto que los datos surgen de un problema relevante para el alumno, tienen para él un significado y tienen que ser interpretados. Por tanto, se aprende mejor qué son los datos reales, y se introducen ideas que no aparecen con los ñdatos inventados por el profesorò: precisión, variabilidad, fiabilidad, posibilidad de medición, sesgo, etc. Se muestra también que la estadística no se reduce a contenidos matemáticos.

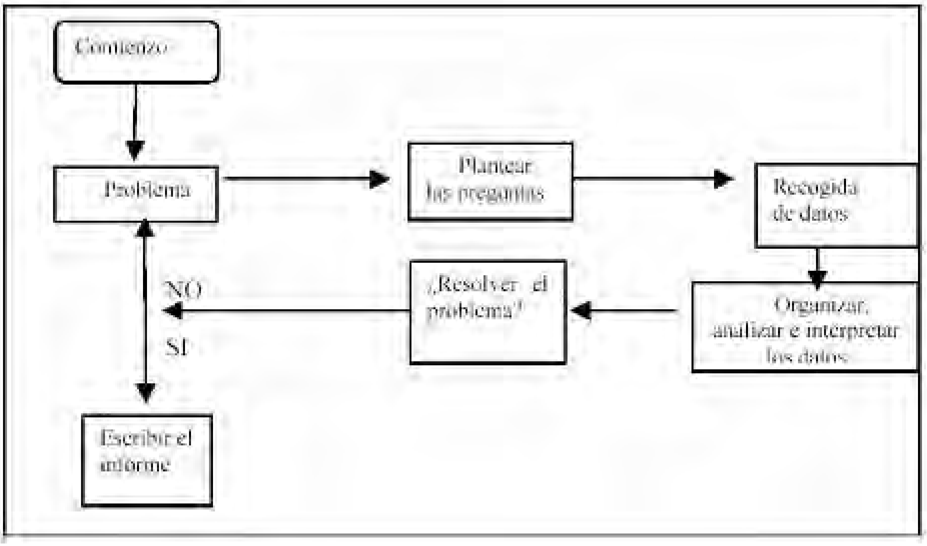
Por otro lado, hay que diferenciar entre conocer y ser capaz de aplicar un conocimiento. La habilidad para aplicar los conocimientos matemáticos es mucho más difícil de lo que se supone, porque requiere no sólo conocimientos técnicos (tales como preparar un gráfico o calcular un promedio), sino también conocimientos estratégicos (saber cuándo hay que usar un concepto o gráfico dado) (Connor, Davies y Payne, 2002). Los problemas y ejercicios de los libros de texto sólo suelen concentrarse en los conocimientos técnicos, mientras que cuando trabajan con proyectos los alumnos desarrollan conocimientos estratégicos (Starkings, 1997).

2.6.2. ESQUEMA DEL TRABAJO EN UN PROYECTO

Los proyectos se conciben como investigaciones sencillas, y abordables por los estudiantes en un tiempo razonable, donde se integra la estadística dentro del proceso general de una investigación (Graham, 1987). La finalidad es que los estudiantes puedan comprender la utilidad de la estadística para resolver problemas de otras materias o de la vida cotidiana. Para conseguir estos objetivos, los proyectos deben escogerse con cuidado, basarse en problemas reales (incluso cuando sean versiones simplificadas de un problema dado) abiertos y apropiados al nivel del alumno. Se comienza planteando un problema práctico y se usa luego la estadística para resolverlo.

La figura 2.1. muestra el esquema de la forma de trabajo en un proyecto en la que se ve que la parte puramente ñmatemáticaò de la estadística (la reducción, análisis e interpretación de los datos) es sólo una de las fases del proyecto, y aún la interpretación ha de hacerse en función del contexto del problema planteado (Pimenta, 2003).

Figura 2.1. Esquema del desarrollo de un Proyecto



Una de las fases más difíciles de un proyecto es el planteamiento de preguntas, pues los alumnos tienen dificultad para formular el problema de forma clara. El papel del profesor es ayudarles a pasar de un tema general a una pregunta que pueda contestarse mediante la estadística. Dicho de otro modo, al pasar del problema en la vida real (que es difuso y no completamente cerrado) a preguntas concretas que puedan resolverse con la estadística, el alumno pasa de la realidad a la matemática y comienza a modelizar.

En algunos casos los datos serán dados por el profesor, pero en otros para completar el proyecto, el alumno necesita recoger datos, que pueden provenir de diversas fuentes: por ejemplo, un experimento en la clase, una encuesta o una medida física. Consideramos importante que el alumno tenga la oportunidad de apreciar la diversidad de los datos estadísticos y experimente estas distintas formas de recogida de datos.

Algunas veces los datos se encuentran disponibles, pero hay que saber localizarlos de diferentes fuentes, como libros o anuarios estadísticos. Internet proporciona en la actualidad datos para cualquier tema por el que los alumnos estén interesados, bien a partir de servidores estadísticos específicos donde los profesores de estadística han puesto sus datos al servicio de la enseñanza, bien recurriendo a organismos oficiales como el INE (Instituto Nacional de Estadística), Eurostat, Unesco u otros. En la Tabla 2.1 mostramos algunos de estos servidores

Tabla 2.1. Algunas fuentes de datos en Internet

Journal of Statistical Education	http://www2.ncsu.edu/ncsu/pams/stat/info/jse/homepage.html
The Chance Database	http://www.geom.umn.edu/docs/education/chance/
The Data and Story Library	http://lib.stat.cmu.edu/DASL/
Census at School Project	http://www.censusatchool.ntu.ac.uk
CAUSE	http://www.causeweb.org/
Exploring Data	http://exploringdata.cqu.edu.au/datasets.htm
Statlib	http://lib.stat.cmu.edu/datasets/

La elección del conjunto de datos es muy importante, pues dependiendo del tipo de datos, podremos aplicar más o menos técnicas estadísticas. Por ejemplo, no podemos hacer un histograma con datos cualitativos y no se recomienda hacer la media con datos ordinales. Es importante que los estudiantes experimenten estos diversos tipos de datos para que adquieran capacidad de seleccionar una técnica estadística adecuada. En nuestro estudio trabajaremos con variables cuantitativas pero que sólo toman valores enteros, por lo cual podremos aplicar todas las técnicas que se han estudiado en primaria y secundaria.

Cuando sea posible, los alumnos pueden usar ordenadores para llevar a cabo sus proyectos, no sólo para el análisis de los datos, sino también para elaborar sus informes con un procesador de texto en el que pueden incorporar los resultados gráficos y numéricos de los programas estadísticos. El proyecto contribuye así a aprender estas herramientas informáticas que son hoy día esenciales. Respecto a los programas estadísticos existe una gran variedad, desde programas profesionales, como SPSS hasta las hojas de cálculo como Excel. Nuestros estudiantes han tenido alguna experiencia con la hoja Excel aunque como el proyecto es sencillo, pueden también trabajar a mano o con calculadora.

2.6.3. DESARROLLO DE COMPETENCIAS BÁSICAS A TRAVÉS DE PROYECTOS

El trabajo con proyectos contribuye a la adquisición de las siguientes competencias básicas recogidas en el Decreto de Enseñanzas Mínimas de Educación Secundaria:

- *Competencia en comunicación lingüística.* Durante el desarrollo del proyecto los alumnos se ejercitan en la comunicación oral y escrita; la representación, interpretación y comprensión de la realidad; la construcción y comunicación del conocimiento y la organización y autorregulación del pensamiento. Además adquieren destrezas y actitudes como formarse un juicio crítico, generar ideas, estructurar el conocimiento, tomar decisiones y disfrutar expresándose tanto de forma oral (exponiendo las conclusiones obtenidas a sus compañeros) como escrita (redactando el informe del proyecto).
- *Competencia matemática.* Puesto que han de utilizar y relacionar números enteros, fraccionarios y decimales, los alumnos aplican operaciones básicas, símbolos, formas de expresión y razonamiento matemático. Utilizan las proporciones, funciones, elementos geométricos y de medición. También se ponen en práctica procesos de reflexión que llevan a la solución de los problemas o a la obtención de información, por medio del reconocimiento de las técnicas apropiadas. Al trabajar con los proyectos, los alumnos integraran el conocimiento matemático con conocimientos de otras disciplinas, ya que como hemos visto anteriormente, la parte puramente ñmatemáticaò es sólo una de las fases de los proyectos.
- *Competencia en el conocimiento y la interacción con el mundo físico.* El trabajo con proyectos posibilita la comprensión de sucesos de la actualidad y sus consecuencias y el análisis de fenómenos sociales desde diversos puntos de vista. Hace también posible identificar preguntas o problemas en la vida diaria o en la actualidad y obtener conclusiones basadas en pruebas, con la finalidad de comprender y tomar decisiones. Procura una habilidad progresiva para poner en práctica los procesos y actitudes propios del análisis sistemático de una tarea y de indagación científica, ya que los proyectos se conciben como auténticas investigaciones.
- *Tratamiento de la información y competencia digital.* En las fases de ñrecogida de datosò y ñorganización, análisis e interpretación de los datosò, se habitúa a los alumnos a buscar, obtener y procesar información para transformarla en conocimiento. Los proyectos contribuyen al aprendizaje del uso de calculadora, ordenadores y software y adquirir destrezas de razonamiento para organizar la información, relacionarla, analizarla, sintetizarla y hacer inferencias y deducciones de distinto nivel de complejidad.

- *Competencia social y ciudadana,* pues se adquieren conocimientos diversos y habilidades complejas que permiten participar, tomar decisiones y responsabilizarse de las elecciones y decisiones adoptadas. Además, se conciencia a los alumnos de la importancia de la estadística en la sociedad actual, implicándose a través de procesos estadísticos en la mejora de la sociedad (participando en los censos, etc.). Por otro lado, los proyectos es aconsejable realizarlos en grupos de 2 o 3 personas, lo cual fomenta la cooperación y la valoración del trabajo de los demás. Finalmente ayuda a tener una actitud crítica y reflexiva en la valoración de la información disponible, contrastándola cuando es necesario, y respetando las normas de conducta acordadas socialmente.
- *Competencia para aprender a aprender,* se ejercita la curiosidad de plantearse preguntas, identificar y manejar las diversas técnicas y estrategias con las que afrontar una misma situación problemática y afrontar la toma de decisiones con la información de la que se dispone. Se ejercitan habilidades para obtener información y para transformar dicha información en conocimientos propios.
- *Autonomía e iniciativa personal.* Es preferible que los proyectos sean planteados por los propios alumnos, fomentando así su capacidad de elegir con criterio propio, de ejercitar su imaginación y de llevar adelante las acciones necesarias para desarrollar las acciones y planes personales. Además en el proyecto el estudiante no depende tanto del profesor, pues tiene libertad para elegir las estrategias de resolución.

2.6.4. JUSTIFICACIÓN DEL USO DE UN PROYECTO EN NUESTRO TRABAJO

Por todos los motivos mencionados es deseable el uso de proyectos en la formación estadística de los futuros profesores, para de este modo introducir en la clase de matemáticas el trabajo cooperativo y una filosofía exploratoria, de acuerdo con las recomendaciones recientes sobre enseñanza de la estadística. Además, puesto que el uso de proyectos se recomienda en la enseñanza de la estadística desde la Educación Primaria, al utilizarlos en la formación de profesores se proporciona a éstos ejemplos de proyectos y pautas sobre la metodología de trabajo con los mismos.

En nuestra investigación se escogió un proyecto concreto que cubriese el contenido elemental de estadística contemplado en los Decretos de Enseñanzas Mínimas para la Educación Primaria. El proyecto además es realista pues plantea un problema en el que mediante la recogida y análisis de los datos se puede contestar a una pregunta de interés para los alumnos. De esta manera la estadística se contextualiza, se muestra su utilizad para la comprobación de conjeturas y el análisis de datos en experimentos y esto ayuda a que los estudiantes aprecien su utilidad en diversas situaciones de la vida real.

En el desarrollo del proyecto, los estudiantes no sólo tienen que realizar cálculos matemáticos o gráficos, sino elaborar un informe en que usen estos resultados para apoyar sus argumentos. En la producción del informe final el estudiante debe situar el análisis de sus datos dentro de un argumento coherente y convincente que apoye sus hipótesis. La comunicación de ideas a partir de tablas y gráficos es especialmente importante en el razonamiento estadístico.

Siguiendo a Murray y Gal (2002) hemos tratado de desarrollar en los estudiantes la comprensión e interpretación de la información estadística, que no sólo requiere conocimiento estadístico o matemático, sino también habilidades lingüísticas, conocimiento del contexto, capacidad para plantear preguntas y una postura crítica ante la información. También Nolan y Speed (1999) resaltan la importancia de desarrollar la capacidad discursiva de los estudiantes, como medio de ampliar sus habilidades de pensamiento crítico.

3. INVESTIGACIONES SOBRE COMPRENSIÓN DE GRÁFICOS

3.1. INTRODUCCIÓN

El lenguaje gráfico tiene un papel esencial en la organización, descripción y análisis de datos, al ser un instrumento de *transnumeración*. Esta es una de las formas básicas de razonamiento estadístico definidas por Wild y Pfannkuch (1999), que consiste en obtener una nueva información, al cambiar de un sistema de representación a otro. Por ejemplo, al pasar de un listado de datos a un histograma, el alumno puede percibir el valor de la moda, que antes no era visible en los datos brutos.

La construcción e interpretación de gráficos estadísticos es también parte importante de la cultura estadística a la que cada vez se dedica más atención y que Gal (2002, pg. 2) define como la unión de dos competencias relacionadas*:*

> *a) Interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos estocásticos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos, y b) discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando sea relevante.* (Gal, 2002, pp. 2-3).

A pesar de esta importancia de los gráficos estadísticos, la investigación en didáctica de la matemática nos alerta que la competencia relacionada con el lenguaje de las gráficas estadísticas no se alcanza en la educación obligatoria (Cazorla, 2002) ni tampoco en la preparación de los futuros profesores de Educación Primaria (Espinel, 2007).

En lo que sigue ampliamos un trabajo de survey sobre comprensión de gráficos (Batanero, Arteaga y Díaz, en prensa) para informar sobre las investigaciones relacionadas con las competencias y niveles de comprensión de gráficos estadísticos, así como errores frecuentes relacionados con los mismos.

3.2. ELEMENTOS Y COMPETENCIAS EN LA LECTURA DE GRÁFICOS ESTADÍSTICOS

Un primer punto investigado por diversos autores es la competencia en la lectura de gráficos. Encontramos gráficos en la prensa diaria, en Internet y también en textos de materias como las ciencias sociales. Sería por tanto necesario que una persona culta fuese capaz de comprender la información expresada en los mismos, aunque esta competencia no es sencilla. Cuando se pide a un estudiante interpretar un gráfico, el estudiante debe realizar la traducción entre lo representado en el gráfico y la realidad. Pero esta traducción requiere conocimientos tanto sobre la realidad representada, como sobre los convenios de construcción del gráfico que a veces el estudiante no posee. Un gráfico queda determinado por los siguientes elementos (Curcio, 1987):

- Las *palabras* que aparecen en el gráfico, como el título del gráfico, las etiquetas de los ejes y de las escalas, y que proporcionan las claves necesarias para comprender el contexto, las variables y las relaciones expresadas en el gráfico.
- El *contenido matemático* subyacente en el gráfico. Por ejemplo los conjuntos numéricos empleados y otros conceptos matemáticos implícitos en el gráfico que el estudiante ha de dominar para interpretarlo, como los de área en un gráfico de sectores, longitud en un gráfico de líneas o sistema de coordenadas cartesianas en un diagrama de dispersión.
- Los *convenios específicos* que se usan en cada tipo de gráfico y que se deben conocer para poder realizar una lectura o construcción correcta. Por ejemplo, el alumno ha de conocer que en un diagrama de sectores, la amplitud del sector es proporcional a la frecuencia. En diagrama de dispersión, cada punto representa un caso y las coordenadas del punto los valores de las dos variables representadas. En algunos gráficos estadísticos estos convenios no son sencillos, como ocurre en el gráfico de la caja, que es muy difícil de interpretar si no se estudia la forma en que se construye e interpreta.

Partiendo del análisis anterior, Friel, Curcio y Bright (2001) identifican los siguientes elementos estructurales de un gráfico estadístico:

- El *título* y las *etiquetas* indican el contenido contextual del gráfico y cuáles son las variables en él representadas. Será importante incluir un título y etiquetas no ambiguos.
- El *marco* del gráfico, que incluye los ejes, escalas, y marcas de referencia en cada eje. Dicho marco proporciona información sobre las unidades de medida de

las magnitudes representadas. Puede haber diferentes tipos de marcos y sistemas de coordenadas (lineales, cartesianas bidimensionales o multidimensionales, polares).

- Los *especificadores* del gráfico son los elementos usados para representar los datos, como los rectángulos (en el histograma) o los puntos (en el diagrama de dispersión). Los autores nos alertan de que no todos los especificadores son igualmente sencillos de comprender, sugiriendo el siguiente orden de dificultad: Posición en una escala homogénea (gráficos de línea, de barras, de puntos, algunos pictogramas e histogramas); posición en una escala no homogénea (gráficos polares, gráficos bivariantes); longitud (gráficos poligonales o estrellados sin ejes de referencia, árboles), ángulo o pendiente (gráfico de sectores, discos), área (círculos, pictogramas), volumen (cubos, algunos mapas estadísticos), color (mapas estadísticos codificados mediante color).

En relación con los anteriores componentes del gráfico, su lectura y construcción, se requieren, según Friel, Curcio y Bright (2001), los siguientes tipos de competencias relacionadas con el lenguaje de los gráficos:

- Reconocer los elementos estructurales del gráfico (ejes, escalas, etiquetas, elementos específicos) y sus relaciones. Esta competencia se adquiere cuando es posible distinguir cada uno de estos elementos y 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 en un gráfico (por ejemplo, ser capaz de predecir como 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. Por ejemplo, cuando un diagrama de dispersión es creciente, comprender que la relación representada entre las dos variables es directa.
- Reconocer cuando 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.

3.3. NIVELES DE COMPRENSIÓN DE GRÁFICOS ESTADÍSTICOS

Además de las competencias anteriores, algunos autores definen niveles de comprensión en la lectura crítica de datos y muestran que no todos los alumnos alcanzan el nivel más alto durante la Educación Secundaria. A continuación resumimos las teorías de diversos autores al respecto.

Bertin (1967) sugiere que la lectura de un gráfico comienza con una *identificación externa,* del tema al que se refiere el gráfico, a través de la comprensión del significado del título y las etiquetas. A continuación se requiere una *identificación interna,* de las dimensiones relevantes de variación en el gráfico: variables representadas y escala. 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. A partir de estos supuestos define diversos niveles de lectura de un gráfico:

- *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 o bien en un diagrama de dispersión leer las coordenadas de uno de los puntos.
- *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. Por ejemplo al determinar visualmente la moda de una distribución en un diagrama de barras, se clasifican los datos en subconjuntos (que tienen un mismo valor para la variable) y se comparan entre si estos subconjuntos para ver cuál tiene mayor frecuencia. Otro ejemplo sería detectar la simetría o asimetría de una distribución a partir de su representación en un histograma.
- *Análisis de la estructura* de los datos, comparando tendencias o agrupamientos y efectuando predicciones. Por ejemplo, cuando se representa en un diagrama de barras adosadas dos distribuciones y se analizan las diferencias en promedios y dispersión de las mismas.

Otra clasificación en niveles de comprensión de los gráficos, muy similar a la anterior, con un gran impacto en educación estadística se debe a Curcio (1989), que denominó a los tres niveles definidos por Bertin como ñ*leer entre los datos*ò (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*" (realizar predicciones e inferencias a partir de los datos sobre informaciones que no se reflejan directamente en el gráfico). Este autor mostró que las principales dificultades aparecen en los dos niveles superiores y que el nivel progresa con la edad de los estudiantes. Friel, Curcio y Bright (2001) amplían la clasificación anterior definiendo un nuevo nivel ñ*leer detrás de los datos* **ò** consistente en valorar críticamente el método de recogida de datos, su validez y fiabilidad, así como las posibilidades de extensión de las conclusiones.

Un modelo algo más complejo es debido a Gerber, Boulton-Lewis y Bruce (1995), quienes diferencian siete niveles de comprensión de gráficos, en función de las competencias de los estudiantes para interpretarlos:

- *Nivel 1.* Los estudiantes no se centran en los datos, sino que asocian algunas características de los mismos, con su conocimiento del mundo en forma imprecisa. Por ejemplo, si les hacemos una pregunta sobre edades de niños representados en un gráfico, pueden responder dando su edad.
- *Niveles 2 y 3.* Los estudiantes se centran en los datos representados, pero de forma incompleta. En el nivel 2 no llegan a apreciar el propósito del gráfico e interpretan sólo aspectos parciales de los datos, por ejemplo, solamente leen una de las barras del diagrama de barras. En el nivel 3 los estudiantes aprecian el propósito del gráfico y analizan todos los elementos uno a uno, pero no llegan a una síntesis global, al no comprender algún elemento específico que es clave en la representación. Un caso sería el estudiante que en una pirámide de población interpreta los grupos de edad (que se refieren a un conjunto de personas) como edades de sujetos individuales.
- *Niveles 4, 5 y 6.* Una vez que el estudiante llega a una síntesis global, puede todavía tener una interpretación estática de los gráficos, y podemos diferenciar tres niveles diferentes. En el nivel 4 los estudiantes son capaces de analizar una a una las variables representadas en el mismo gráfico, pero no conjuntamente, por ejemplo, si representamos la esperanza de vida de hombre y mujeres en diversos países en un gráfico de líneas, los alumnos interpretan por un lado la esperanza de vida de los hombres y por otro los de las mujeres. En el nivel 5 se comparan varias variables representadas en el mismo gráfico; en el ejemplo anterior podrían deducir que la esperanza de vida en las mujeres es superior a la de los

hombres en la mayoría de países. En el nivel 6 los estudiantes usan los gráficos para apoyar o refutar sus teorías. No sólo comparan varias variables en el mismo gráfico, sino sacan conclusiones generales respecto a una hipótesis, por ejemplo, podrían usar el gráfico anterior para refutar la idea de que la mujer es más débil que el hombre.

- *Nivel 7.* En el último nivel los estudiantes son capaces de hacer extrapolaciones, y hacer predicciones para otros datos no representados en el gráfico; en el ejemplo anterior, podrían deducir la esperanza de vida del hombre, conocida la esperanza de vida de la mujer, para un determinado país no representado en el gráfico.

Cuando estos niveles se aplican, no sólo a la interpretación de los gráficos sino a su valoración crítica, los niveles superiores se modifican ligeramente (Aoyama y Stephen, 2003, Aoyama, 2007). Supongamos, por ejemplo, que se da a los estudiantes un gráfico que presenta datos sobre el número de horas que los adolescentes dedican a los juegos con videoconsola y el número de episodios de violencia escolar en que se ven implicados. La gráfica muestra claramente un crecimiento del número de episodios de violencia cuando aumenta el tiempo dedicado a este tipo de juegos. Se pregunta a los estudiantes si piensan que la violencia escolar disminuiría si se prohibiesen las videoconsolas. Una vez que los estudiantes llegan a la fase superior en la clasificación anterior, todavía podríamos diferenciar tres grupos, en función de su capacidad crítica, respecto a la información reflejada en el gráfico.

- *Nivel Racional/Literal.* Los estudiantes leen correctamente el gráfico, incluyendo la interpolación, detección de tendencias y predicción. Para responder la pregunta planteada, usan las características del gráfico, pero no cuestionan la información, ni dan explicaciones alternativas. Una respuesta típica sería ñSí, ya que el grupo de chicos que jugó a los juegos durante mucho tiempo también tuvo muchos episodios de violencia.ò
- *Nivel Crítico.* Los estudiantes leen los gráficos, comprenden el contexto y evalúan la fiabilidad de la información, cuestionándola a veces, pero no son capaces de buscar otras hipótesis: ñPienso que no, pues aunque los que más juegan aparecen como más violentos en el gráfico, podría haber otras causas, aunque no me imagino cuálesò.

- *Nivel Hipotético:* Los estudiantes leen los gráficos, los interpretan y evalúan la información. Forman sus propias hipótesis y modelos: ñNo estoy de acuerdo en que la causa de la violencia sea el juego, quizás la falta de atención de los padres puede llevar a la vez a que el chico sea violento y juegue más con la consolaò.

3.4. ERRORES EN LA LECTURA Y CONSTRUCCIÓN DE GRÁFICOS

Además de las capacidades de lectura de los gráficos, otras investigaciones analizan los errores frecuentes en la producción de los mismos. El primer paso en su construcción sería elegir un gráfico adecuado, tanto al tipo de variable, como al problema planteado, pero los estudiantes fallan con frecuencia en esta elección. Li y Shen, (1992) analizaron los gráficos en los proyectos estadísticos de sus estudiantes, encontrado alumnos que utilizan polígonos de frecuencias con variables cualitativas, o diagrama de barras horizontal para representar datos que debieran representarse en un diagrama de dispersión. Otras veces, construyen gráficos sin sentido, por ejemplo se representan variables no relacionadas entre sí en un mismo gráfico.

Respecto a las escalas de los gráficos construidos por los estudiantes Li y Shen (1992) encontraron los siguientes problemas:

- Elegir una escala inadecuada para el objetivo pretendido (por ejemplo no se cubre todo el campo de variación de la variable representada)
- Omitir las escalas en alguno de los ejes horizontal o vertical, o en ambos.
- No especificar el origen de coordenadas.
- No proporcionar suficientes divisiones en las escalas de los ejes.

Encontramos también investigaciones sobre errores en la lectura y comprensión de gráficos específicos. En el diagrama de barras, al variar la disposición de los datos (por ejemplo al usar barras horizontales en lugar de verticales) los estudiantes pueden tener errores simples de lectura (Pereira-Mendoza y Mellor, 1990). Lee y Meletiou (2003) nos alertan de cuatro principales categorías de razonamientos erróneos a la hora de construir, interpretar y aplicar los histogramas en diferentes contextos de la vida real:

- Percepción de los histogramas como representación de datos aislados, suponiendo que cada rectángulo se refiere a una observación particular y no a un intervalo de valores.

- Tendencia a observar el eje vertical y comparar las diferencias en las alturas de las barras cuando comparan la variación de dos histogramas.
- Interpretación determinista, sin apreciar que los datos representan un fenómeno aleatorio que podría variar al tomar diferentes muestras de la misma población.
- Tendencia a interpretar los histogramas como gráficos de dos variables (es decir, como diagramas de dispersión).

Respecto a los gráficos de caja, aunque su estudio se recomienda en la Educación Secudaria Obligatoria, Bakker, Biehler y Konold (2004) indican que esta representación no permite a los estudiantes percibir los valores individuales de los datos y además son muy diferentes a otros gráficos usados por los alumnos ya que están basados en la mediana y cuartiles, conceptos que no son intuitivos para los alumnos.

El ordenador en ocasiones contribuye a empeorar los problemas de los estudiantes. Ben-Zvi y Friedlander (1997) analizan los gráficos producidos por sus alumnos al trabajar con proyectos de análisis de datos con ayuda del ordenador, identificando cuatro categorías:

- *Uso acrítico:* Los estudiantes construyen gráficos rutinariamente aceptando las opciones por defecto del software, aunque no sean adecuadas. Tienen también dificultad en valorar las relaciones sugeridas en sus representaciones gráficas, identificando sólo la información obvia, por ejemplo los máximos.
- *Uso significativo de una representación:* Los estudiantes construyen correctamente un gráfico si se les indica cuál ha de utilizar; también lo pueden justificar en base al tipo de datos o al problema planteado. Son capaces de modificar y transformar la gráfica (por ejemplo, cambiar una opción del software) e interpretan los resultados, pero no son capaces de seleccionar la gráfica más adecuada cuando tienen varias para elegir.
- *Manejo significativo de representaciones múltiples:* En este caso, los alumnos toman decisiones en la selección de los gráficos más adecuados, toman en consideración cuál es la contribución de éstos a su problema.
- *Uso creativo:* Cuando el alumno elabora un gráfico correcto, no habitual, para presentar y justificar sus ideas.

3.5. COMPRENSIÓN GRÁFICA DE LOS FUTUROS PROFESORES

Las dificultades con los gráficos estadísticos no son exclusiva de los estudiantes, sino que también se presentan en los futuros profesores, más concretamente en los alumnos de las Facultades de Ciencias de la Educación. Bruno y Espinel (2005) estudian la forma en que futuros profesores construyen un histograma de frecuencias a partir de una lista de datos. Aproximadamente la mitad de los participantes en su investigación tuvieron errores al representar los intervalos de variación de la variable en el eje *X*, omitir intervalos de frecuencia nula, o usar rectángulos no adosados, a pesar de que la variable fuese continua. En cuanto al polígono de frecuencias, los errores más frecuentes en la investigación citada fueron no unir por las marcas de clase, omitir el intervalo de frecuencia nula en el polígono y confundir la frecuencia y valor de la variable.

Continuando la investigación, Bruno y Espinel tratan de comparar los errores de los futuros profesores en la construcción del histograma y el polígono de frecuencias, con su evaluación de histogramas producidos por posibles estudiantes. Prácticamente todos los estudiantes cometen algún error al construir los gráficos, pero lo más preocupante es la falta de coherencia entre la forma que los futuros profesores construyen un gráfico y la forma en que lo evalúan. Además, en caso de coherencia se trata de alumnos que cometen errores en la interpretación de los gráficos, y también consideran correctos los gráficos incorrectos de sus posibles estudiantes.

Preocupadas por estos resultados las autoras continúan la investigación utilizando un cuestionario sobre cultura y razonamiento estadístico en relación a los gráficos y comparando sus resultados con los de otros estudiantes universitarios americanos (Espinel, 2007; Espinel, Bruno y Plasencia, 2008). Para ello dieron a 190 estudiantes de Magisterio un conjunto de ítems de opciones múltiples en los que se evaluaba su competencia gráfica y que habían sido tomados del proyecto ARTIST (delMas, Garfield y Ooms, 2005), donde se pasó a 350 estudiantes de primer curso de universidad. Aunque en ambos colectivos las tareas fueron difíciles, la dificultad fue mayor para los futuros profesores españoles, en particular al tratar de predecir la forma de un gráfico a partir de la descripción verbal de variables conocidas por los estudiantes o en la lectura de histogramas.

Monteiro y Ainley (2006, 2007) indican que la lectura de los gráficos usuales dentro del contexto escolar es una tarea más limitada que la posible interpretación de dichos gráficos fuera de dicho contexto, porque, mientras en la escuela sólo pedimos a los estudiantes una respuesta correcta desde el punto de vista matemático, fuera de la escuela interviene también el conocimiento del contexto en el que se sitúa el tema del gráfico. Al estudiar la competencia de futuros profesores en la lectura de gráficos tomados de la prensa diaria, en primer lugar encontraron que muchos no tenían conocimientos matemáticos suficientes para llevar a cabo dicha lectura. La mayoría indicó que no tuvo formación específica en la lectura de gráficos estadísticos durante sus estudios en la universidad y reconoció su necesidad que tenían de formación al respecto. También se observó como 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 sobre tiempo de gestación de diferentes especies animales.

Tampoco es bueno el conocimiento didáctico de los futuros profesores respecto a la enseñanza de los gráficos. En un estudio cualitativo González y Pinto (2008) analizan la forma en que algunos futuros profesores seleccionan y clasifican gráficos para preparar una unidad didáctica dirigida a la enseñanza del tema en educación primaria. Los estudiantes parecen desconocer todo lo relacionado con las dificultades de los futuros alumnos o los niveles de comprensión de gráficos. Además, a la hora de clasificarlos o seleccionarlos sólo tienen en cuenta los aspectos procedimentales y no la utilidad del gráfico para un problema particular o las posibles actividades interpretativas que pueden llevarse a cabo con ellos.

3.6. CONCLUSIONES SOBRE LAS INVESTIGACIONES PREVIAS

La investigación reseñada muestra que la lectura e interpretación del lenguaje gráfico es una habilidad altamente compleja, que no se adquiere espontáneamente, pero, por desgracia, tampoco parece alcanzarse con la enseñanza. Una posible explicación de las dificultades surgidas con el lenguaje gráfico es que la simplicidad de este es sólo aparente, pues incluso el más simple de los gráficos puede considerarse un modelo matemático. Al reducir los datos, pasando de casos individuales a los valores de una variable y sus frecuencias, se introduce la *distribución de frecuencias*, concepto complejo, que se refiere al agregado (población o muestra) y no a los datos particulares.

Los gráficos estadísticos se encuentran presentes en la vida cotidiana, tanto en los medios de información e Internet, como en los textos escolares de diferentes materias y en el trabajo profesional. Por ello, un ciudadano competente debiera poder realizar una lectura crítica de dichos gráficos, identificar las tendencias y variabilidad de los datos, y detectar los posibles errores que puedan distorsionar la información representada (Schield, 2006).

Asimismo debiera conocer los convenios de construcción de los diferentes tipos de gráficos y ser capaz de construir correctamente un gráfico sencillo. Esto es todavía más necesario en los futuros profesores responsables de formar a los niños en el tema. Sin embargo la investigación reseñada indica que los futuros profesores de Educación Primaria tienen dificultades con el lenguaje gráfico que han de transmitir a sus alumnos y han de utilizar como herramienta en su vida profesional. Nuestra investigación trata de contribuir a proporcionar más información al respecto con el fin de poder mejorar en el futuro la preparación de los profesores.

4. EL ESTUDIO EMPÍRICO

4.1. INTRODUCCIÓN

En este capítulo presentamos una investigación, donde analizamos los gráficos producidos por una muestra de estudiantes de Magisterio. Aunque la actividad con la que se tomaron los datos es más amplia, y requería también la elaboración de tablas y resúmenes estadísticos, nosotros nos centramos sólo en el estudio de los gráficos producidos, que analizaremos desde distintos puntos de vista.

Primeramente describimos la muestra y el contexto en que se realizó la actividad, así como la enseñanza de estadística que los alumnos participantes habían tenido y realizamos un análisis a priori del proyecto con el cual se tomaron los datos para realizar el estudio empírico. Seguidamente clasificamos los gráficos construidos por los estudiantes según unos niveles de complejidad semiótica definidos teniendo en cuenta algunas ideas teóricas de Font, Godino y DôAmore (2007). Para cada uno de los niveles presentaremos un ejemplo de gráfico correcto y otro de gráfico incorrecto, realizando un análisis semiótico de los mismos.

Analizamos también la corrección del gráfico, su interpretación y las conclusiones obtenidas en función de su complejidad semiótica y finalizamos con la discusión de las conclusiones.

4.2. DESCRIPCIÓN DE LA MUESTRA

La actividad se ha llevado a cabo con 93 alumnos que cursaban el segundo año del plan de estudio de la Diplomatura de Magisterio, en la especialidad de Educación Primaria. Los alumnos estaban divididos en tres grupos, todos ellos de la asignatura ñCurrículo de matemáticas en Educación Primariaò, impartido cada uno por diferente profesor, aunque la actividad, el material usado y el análisis es el mismo para todos los grupos.

Como ya se ha visto en el punto 2.5, los estudiantes de nuestra muestra ya habían cursado la asignatura de ñMatemáticas y su Didácticaò en el primer curso de la Diplomatura de Magisterio, aunque no todos la aprobaron.

Los alumnos de nuestra muestra en la asignatura ñMatemáticas y su Didácticaò, así como en Bachillerato, Educación Primaria y Secundaria habían estudiado los gráficos y resúmenes estadísticos elementales.

4.3. ANÁLISIS A PRIORI DEL PROYECTO

4.3.1. LA ACTIVIDAD

La actividad con la que se tomaron los datos era parte de una práctica a la que se dedicó dos sesiones de clase de dos horas de duración cada una, dentro de la asignatura citada (Currículo de Matemáticas en Educación Primaria) y se relaciona con el cuarto bloque del currículo de matemáticas en la Educación Primaria: *Tratamiento de la información, azar y probabilidad.*

Dicha práctica, titulada *Evaluación de experiencias de enseñanza: estadística en primaria* tenía como objetivos adquirir competencias de análisis didáctico de experiencias de enseñanza y también competencias de búsqueda de recursos didácticos. Puesto que es difícil llevar a todos los estudiantes a un aula de primaria a observar una sesión de enseñanza, en lugar de ello, se realiza una experiencia de enseñanza de la estadística en el mismo aula, siendo los alumnos los estudiantes para profesor (primera sesión) y posteriormente se analiza desde el punto de vista didáctico la experiencia realizada (segunda sesión).

Nosotros nos centramos sólo en la experiencia llevada a cabo en la primera de estas sesiones. En ella se propuso a los futuros profesores la realización de un proyecto de análisis de datos, titulado ñComprueba tus intuiciones sobre el azarò, en el que ellos mismos obtuvieron los datos necesarios. Al finalizar la sesión se dio a los estudiantes una copia con los datos obtenidos en la clase y se les pidió producir individualmente un informe escrito en el cual debían responder la pregunta que se les planteaba. Los estudiantes tuvieron libertad para elegir los gráficos o resúmenes estadísticos que considerasen convenientes e incluso usar ordenadores. A la semana siguiente, en la segunda sesión, se recogieron y analizaron las producciones de los estudiantes.

4.3.2. EL PROYECTO

El proyecto ñComprueba tus intuiciones sobre el azarò es parte de una unidad didáctica diseñada para introducir los temas sobre ñtratamiento de la información, azar y probabilidadò en los primeros años de Educación Secundaria Obligatoria, pero su contenido estadístico es también adecuado para la formación de maestros (Batanero, 2001). Fines complementarios de este proyecto son mostrar la utilidad de la estadística para probar conjeturas y comprobar que las intuiciones sobre el azar a veces son engañosas. La secuencia de actividades se describe a continuación (una descripción más completa puede analizarse en Godino, Batanero, Roa y Wilhelmi, 2008).

1. Presentación del proyecto, instrucciones iniciales y discusión colectiva. Se comenzó la sesión con una discusión sobre si las personas tienen o no buena intuición respecto a los fenómenos aleatorios. Conseguido el interés de los alumnos en el tema, se propuso a los futuros profesores llevar a cabo un experimento para decidir si la clase en conjunto tenía o no buenas intuiciones respecto a dichos fenómenos. El experimento consiste en inventar una secuencia de 20 posibles resultados al lanzar una moneda equilibrada (sin lanzarla realmente) de modo que la secuencia pueda pasar como aleatoria para otra persona y comparar con los resultados de 20 lanzamientos reales de una moneda. Este experimento está adaptado de otros realizados en las investigaciones sobre percepción subjetiva de la aleatoriedad (por ejemplo, de la investigación de Serrano, 1996).

Figura 4.1. Hoja de registro

Secuencia simulada																			
Secuencia real																			

2. Experimentos individuales y recogida de datos. Los futuros profesores realizaron individualmente el experimento, inventando cada uno de ellos una secuencia de 20 lanzamientos (secuencia simulada) y anotando los resultados en una hoja de registro (Figura 4.1) escribiendo C para cara y + para cruz. A continuación realizaron cada uno 20 lanzamientos de una moneda, anotando los resultados en la segunda parte de la hoja (secuencia real).

3. Discusión y actividades de la clase. Finalizado el experimento, el profesor inicia la discusión sobre cómo comparar los resultados de la totalidad de la clase en las secuencias reales y simuladas. Entre otras variables, algunos estudiantes sugirieron comparar el número de caras en las dos secuencias, que son las variables que consideramos en este trabajo. Al acabar la clase el profesor proporcionó una hoja de registro que los alumnos rellenaron con el número de caras en las secuencias real y simulada, obtenidos por cada uno de los estudiantes del grupo en sus experimentos individuales.

Observamos que en esta actividad aparecen dos variables aleatorias y sus correspondientes variables estadísticas:

- La variable aleatoria ξ_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, sería igual a $np=10$ y su varianza igual a $npq=5$.
- La variable aleatoria ξ_s: número de caras en la 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 realiza 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 ξ_r. El estudiante ha de usar esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria ξ_r puesto que sus conocimientos matemáticos no le permiten trabajar directamente con ξ_r.
- La variable estadística Y_s o resultados de una muestra de *m* valores de la variable aleatoria ξ_s. El estudiante ha de usar esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria ξ_s puesto que sus conocimientos matemáticos no le permiten trabajar directamente con ξ_s.
- Más concretamente, el estudiante ha de comparar las características de las distribuciones de las variables Y_r *y* Y_s para conjeturar sobre las semejanzas o diferencias entre ξ_r *y* ξ_s y de acá deducir sobre las intuiciones respecto al azar.

4. Elaboración de un informe individual. Una vez que los estudiantes tienen una hoja con los datos de estas variables, han de resumirlos, formando la distribución de frecuencias de las correspondientes variables estadísticas para proceder, primeramente al análisis de cada una de estas dos variables estadísticas y luego a la comparación de las principales diferencias en su distribución. De ello esperamos extrapole a las diferencias entre las correspondientes variables aleatorias y de ahí obtenga conclusiones respecto a las intuiciones de la clase en su conjunto sobre la aleatoriedad. Esperamos que los estudiantes se den cuenta de que las variables tienen la misma media, pero distinta dispersión. Las intuiciones sobre el número de caras son correctas respecto al promedio, pero hay una mala intuición respecto a la dispersión, pues la secuencia de los alumnos tiene menos variabilidad que las secuencias reales. En otras palabras no se percibe la variabilidad real de un proceso aleatorio, aunque si sus tendencias (Serrano, 1996).

4.3.3. SOLUCIÓN ESPERADA DE LOS ESTUDIANTES

Como hemos dicho, el profesor dio a los alumnos una hoja de registro como la que reproducimos en la Tabla 4.1, que muestra los datos obtenidos en uno de los grupos de estudiantes que participaron en este Proyecto en otras asignaturas.

Tabla 4.1. Número de caras obtenidas en cada secuencia en uno de los grupos

Alumno	Secuencia real	Secuencia simulada
1	10	11
2	12	11
3	11	11
4	10	8
5	11	7
6	9	8
7	10	9
8	11	11
9	9	10
10	10	9
11	10	9
12	10	9
13	7	9
14	10	14
15	10	7
16	10	10
17	10	9
18	12	10
19	11	11
20	10	13
21	9	11
22	10	8
23	10	8
24	9	11
25	10	12
26	12	9
27	11	8

Para recoger estos datos, una vez finalizados los experimentos, el profesor pide a los alumnos que, según el orden en que están colocados en la clase, vayan dando en voz alta sus datos para el número de caras en cada una de las secuencias. Cada vez que un estudiante da sus datos, todos los compañeros lo anotan en una nueva fila de la tabla. Como vemos se tiene una tabla de datos con tantas filas como alumnos en la clase y dos variables estadísticas: número de caras en la secuencia real y simulada.

Al comenzar el trabajo, disponiendo ya los alumnos de la lista de datos, se ha pasado de un problema extra-matemático (decidir sobre las intuiciones respecto al azar del conjunto de la clase) a otro matemático (comparar dos listas de datos). Para resolver el problema planteado, esperamos que los estudiantes identifiquen el valor máximo y mínimo del número de caras en las secuencias real y simulada y preparen una tabla de frecuencias como la Tabla 4.2 para resumir la información, donde se recoja la frecuencia absoluta, porcentaje, quizás también la frecuencia acumulada. Sería necesaria también otra tabla similar para resumir los datos de la secuencia real.

Se espera también que los alumnos calculen resúmenes estadísticos tales como las medidas de posición central: media, mediana y moda y de dispersión (al menos el rango) que se han estudiado el curso anterior y cuyos valores para el conjunto de datos de la Tabla 4.1 se muestran en la Tabla 4.3.

Tabla 4.2. Número de caras en la secuencia simulada

Número de caras	Frecuencia	F. Acumulada	Porcentaje (%)
7	1	1	3,7
8	0	1	0
9	4	5	14,8
10	14	19	51,85
11	5	24	18,5
12	3	27	11,1
Total	27	27	100

Tabla 4.3. Medidas de posición central y de dispersión en las secuencias

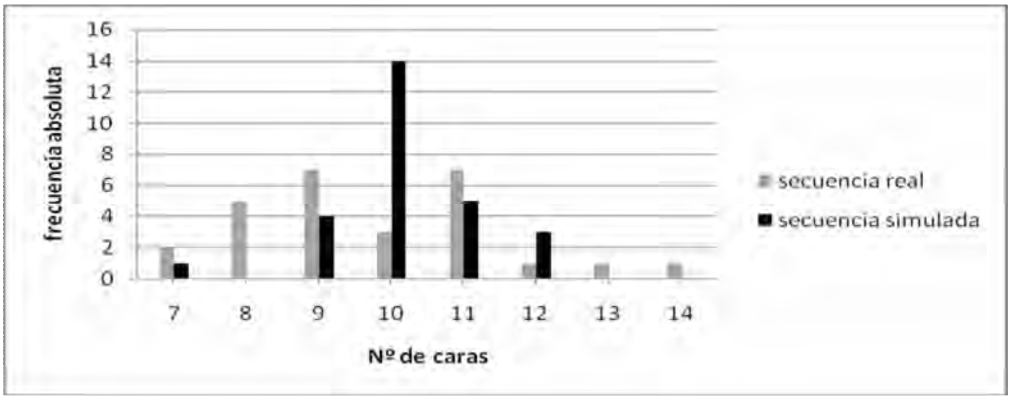
	Media	Mediana	Moda	Rango	Desviación típica
Secuencia simulada	10,14	10	10	5	1,06
Secuencia real	9,74	9	9 y 11 (bimodal)	7	1,74

Lo deseable sería que realizasen gráficos simultáneos para las dos distribuciones, ya que esto facilitaría la comparación de ambas y como consecuencia la mejor interpretación de la información estadística puesta en juego y la obtención de conclusiones en relación a las intuiciones del conjunto de la clase sobre el azar. En las siguientes figuras (4.2, 4.3, 4.4) se muestran ejemplos de algunos gráficos de este tipo que podrían utilizar los alumnos. Aunque todos ellos representan la misma información, los objetos matemáticos puestos en juego varían. En el gráfico de puntos, por ejemplo, no hay un eje que represente la frecuencia de los valores de la variable, sino que esta frecuencia se representa por el número de cruces en el gráfico asociado a cada valor de las variables. Dicha frecuencia está representada por la altura de la barra en el gráfico de barras y por la coordenada *Y* del punto en el gráfico de líneas, que es donde verdaderamente aparece una representación en coordenadas cartesianas.

Figura 4.2. Gráfico de puntos conjunto para la secuencia real y simulada

Secuencia simulada	Nº. Caras	***Secuencia real***
X	7	XX
	8	XXXXX
XXXX	9	XXXXXXX
XXXXXXXXXXXXXX	10	XXX
XXXXX	11	XXXXXXX
XXX	12	X
	13	X
	14	X

Figura 4.3 Gráficos de barras adosados para el nº de caras para ambas secuencias



Siendo conscientes de que este tipo de gráficos, al estar representados en ellos dos variables, son de mayor complejidad para los alumnos, podemos admitir también que los estudiantes hagan gráficos separados para la distribución de cada secuencia, siempre que se utilice el mismo tipo de gráfico para ambas y la misma escala en los dos gráficos, para que así se facilite la comparación de las distribuciones en estudio. En la Figura 4.5 se muestra un gráfico de barras para cada una de las secuencias, en los cuales se ha utilizado la misma escala. Tanto en el gráfico de barras como en el de líneas se puede visualizar mejor la moda, la dispersión y la forma que tienen las distribuciones, que en otros gráficos como los de sectores, que aunque estén correctamente construidos, es más difícil extraer de ellos diferencias importantes entre las dos distribuciones que indiquen si nuestra intuición sobre el azar nos engaña.

Figura 4.4. Gráfico de líneas adosado del nº de caras para ambas secuencias

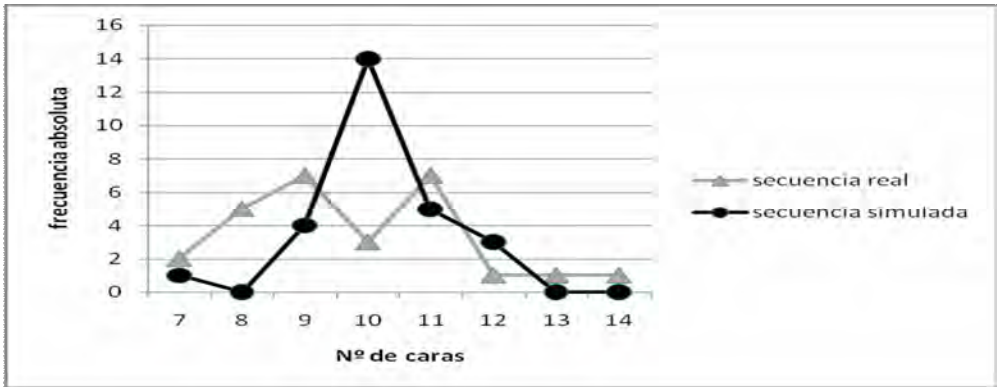
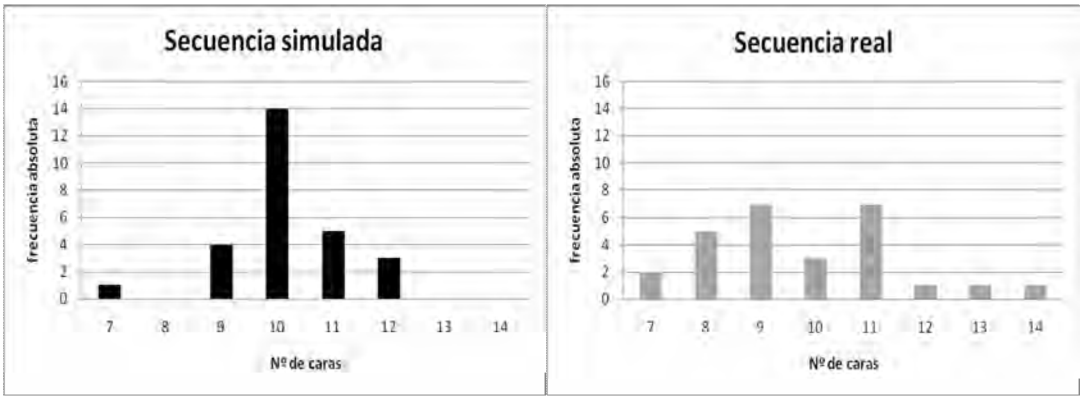


Figura 4.5. Gráficos de barras secuencia simulada y secuencia real



Interpretación y conclusiones

Una característica del número de caras en una secuencia real es que, en general, es más variable de lo que nuestra intuición nos sugiere como se vio en la investigación de Serrano (1996) con alumnos de bachillerato y futuros profesores. Mientras que los valores medios del número de caras coinciden, aproximadamente en las secuencias reales y simuladas de los participantes en dicho estudio, la dispersión fue mucho mayor en las secuencias reales. El autor indica que somos muy exactos al reflejar la equiprobabilidad de resultados, incluso más exactos de lo debido, puesto que la secuencia simulada tiene menos dispersión que la real.

Al estudiar los diferentes gráficos (4.2 a 4.5) y lo resúmenes estadísticos presentados la Tabla 4.3, se observa que hemos obtenido una distribución bimodal lo cual sugiere la necesidad de usar la media o la mediana para llevar a cabo la comparación: $\bar{x}$=10,14 para las secuencias simuladas $\bar{x}$=9,74 para las reales en nuestro ejemplo, donde vemos que los valores son muy parecidos entre sí y casi iguales al valor teórico np=10 de la variable aleatoria número de caras en 20 lanzamientos de una moneda equilibrada, que es una variable aleatoria Binomial. Las medianas son respectivamente iguales a 10 y 9. Por tanto los estudiantes han reproducido intuitivamente los promedios de la variable ñnúmero de cara en 20 lanzamientos de una monedaò.

También surge en esta actividad la idea de dispersión de una forma sencilla. Bien a través del recorrido o del 50 % de casos centrales se observa mayor dispersión en la secuencia real, donde el 50% de los casos centrales se presentan en el intervalo (9-11) y el recorrido es 7, mientras que en la secuencia simulada el 50% de casos centrales se reduce al valor 10 y el recorrido es 5. Esperamos que los estudiantes, bien calculando los estadísticos, bien a partir de los gráficos perciban la igualdad aproximada de las medidas de posición central y la mayor dispersión de las secuencia reales.

La conclusión respecto a las intuiciones es que los alumnos de los grupos donde se realizó el experimento tienen una buena percepción del valor esperado del número de caras en 20 lanzamientos, puesto que la mayoría produce exactamente 10 caras. Sin embargo, la variabilidad del número de caras no se percibe, suponiendo mayor regularidad que la existente en un proceso aleatorio.

4.4. GRÁFICOS CONSTRUIDOS POR LOS ESTUDIANTES

Una vez recogido el informe de análisis de datos del proyecto que entregó cada estudiante, se clasificaron los gráficos incluidos como parte del análisis, desde varios puntos de vista. En este trabajo nos limitamos a analizar los gráficos relacionados con el número de caras, aunque en la actividad también se analizaron otros dos pares de variables (número de rachas y longitud de la racha más larga en las secuencias reales y simuladas). Sin embargo, los resultados se pueden generalizar pues los alumnos se limitan a repetir el mismo análisis para los tres pares de variables. Por ejemplo, si sólo hacen tablas de frecuencias, producen estas con las mismas columnas para las seis variables del proyecto.

4.4.1. TIPOS DE GRÁFICOS USADOS Y COMPLEJIDAD SEMIÓTICA

En primer lugar tuvimos en cuenta los datos y número de variables representadas, para definir un nivel de complejidad semiótica del gráfico, usando algunas ideas teóricas de

Font, Godino y DôAmore (2007). El propósito es definir una jerarquía en la construcción de gráficos estadísticos, puesto que no hemos visto recogido este tipo de jerarquía en las investigaciones previas, mientras que son los muchos los autores que definen niveles de lectura o comprensión de los gráficos (por ejemplo Curcio, 1987, 1989; Friel, Curcio y Bright, 2001). Esta categorización la hemos presentado en Arteaga, Batanero, y Ruiz (en prensa) y Arteaga, Ortiz y Batanero (2008).

Como ya hemos visto en el apartado 2.2, Font, Godino y DôAmore estudian los objetos matemáticos que surgen de las prácticas matemáticas realizadas para resolver un problema con la finalidad de permitir analizar con detalle los procesos didácticos. Para ello clasifican dichos objetos e introducen toda una ontología de los objetos matemáticos que están implícitos en las prácticas matemáticas. En nuestro estudio nos serán de utilidad estas ideas teóricas además de otras como función semiótica y configuraciones cognitivas (ya descritas en el apartado 2.2).

Así, en la actividad proponemos un problema (comparar las dos distribuciones dadas para decidir sobre las intuiciones del conjunto de alumnos) y analizamos las prácticas realizadas para resolverlo. Específicamente, el gráfico producido involucra una serie de acciones, conceptos y propiedades que varían en los diferentes gráficos. Por tanto, varían las funciones semióticas subyacentes en su construcción e interpretación por parte de los estudiantes, que incluye también el relacionar la pregunta inicial con los gráficos, por medio de una argumentación.

En consecuencia, 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 Font, Godino y DôAmore (2007). Usando estas ideas, en nuestro trabajo hemos realizado un análisis semiótico de los gráficos producidos por los estudiantes, definiendo los niveles de complejidad semiótica que se describen a continuación:

No produce gráficos. Un total de cinco estudiantes en la muestra (n=93) se limitaron a realizar cálculos estadísticos, en la mayoría de los casos reducidos a tablas o medidas de posición central (media, mediana y/o moda); a veces también la dispersión (rango, desviación típica). Una parte de los que no producen gráficos se limita a presentar dichos resúmenes sin interpretarlos y el resto obtiene una conclusión sobre las intuiciones a partir de la comparación de las medidas de posición central y en algunos casos de las de dispersión en ambas secuencias. No tendremos en cuenta estos estudiantes en nuestro análisis, pues nuestro interés se centra en los gráficos estadísticos producidos por los estudiantes.

C1. *Representa sólo sus resultados individuales.* Algunos alumnos producen una gráfica para representar los datos obtenidos en su experimento particular, sin considerar los datos 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. Generalmente manifiestan una intuición errónea sobre la aleatoriedad suponiendo que una buena intuición implicaría que su secuencia simulada fuese idéntica en alguna característica a su secuencia real.

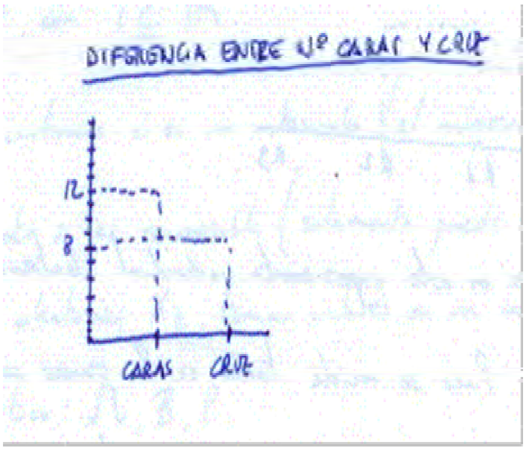
Aunque la impredecibilidad es una diferencia esencial entre experimentos aleatorios y deterministas, estos estudiantes tratan de predecir el resultado o al menos predecir una parte de los resultados de la secuencia aleatoria. Este fenómeno, denominado *ilusión de control* fue observado originalmente 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.

Estos estudiantes representan la frecuencia de resultados obtenidos en sus propias secuencias; en el caso de la secuencia real, consideran solamente la variable aleatoria ñresultado del lanzamiento de una moneda equilibradaò, es decir, una variable Bernoulli que puede tomar los siguientes valores ñCaraò o ñCruzò. Así estos alumnos llegan a una idea intuitiva de variable estadística y su distribución, siendo dicha variable los resultados de los 20 lanzamientos de una moneda que cada uno de los alumnos realizan, es decir se trata de una muestra de 20 valores de la variable aleatoria Bernoulli.

En la Figura 4.6 se muestra un ejemplo correcto. El gráfico representa las frecuencias de caras y cruces en los 20 lanzamientos del estudiante. En este caso concreto, el estudiante usa un diagrama que podríamos interpretar como de barras, aunque sustituye las barras por líneas de puntos. También traza una línea punteada, no necesaria, para marcar las frecuencias que corresponden a cada resultado.

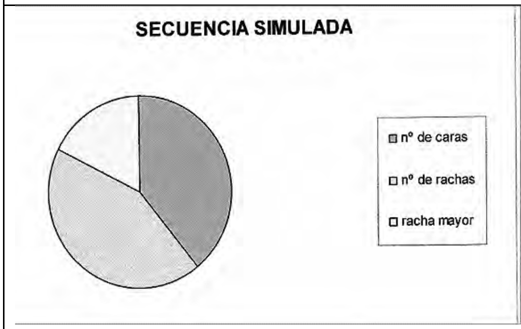
El título del gráfico es confuso ya que lo que se representa son las frecuencias de cada uno de los resultados y no las diferencias. Los ejes no tienen rótulos aunque sí etiquetas para los valores de las variables representadas en los mismos. En el eje *X* el significado de estas etiquetas es claro, sin embargo en el eje Y debemos interpretar que lo que se representa son frecuencias absolutas. Para realizar el gráfico, el alumno tiene que haber hecho un recuento del número de caras y cruces, considerando el experimento aleatorio, sus posibles resultados, la variable estadística y las correspondientes frecuencias. Usa implícitamente la proporcionalidad en la escala del eje *Y*. El orden de la escala en el eje *X* es artificial.

Figura 4.6. Análisis semiótico de un gráfico correcto de la categoría C1

Representación de resultados en la secuencia del estudiante	*Configuración cognitiva*
 Lenguaje - Grafico: líneas, puntos, marcas - Verbal: título, etiquetas - Numérico: para representar las frecuencias de las caras y de las cruces en el experimento en el eje *Y*. - No hay rótulos en los ejes. - Rótulo confuso para el título del gráfico.	Conceptos y proposiciones - Experimento aleatorio: lanzar una moneda. - Sucesos o resultados del experimento, espacio muestral. - Variable aleatoria que toma dos valores: ñCaraò y ñCruzò (Bernoulli) - Variable estadística: resultados de repetir 20 veces el experimento. - Frecuencias absolutas de caras y cruces. - Números naturales. - Puntos y segmentos. - Perpendicularidad y paralelismo. - Proporcionalidad. - Sistema cartesiano de representación, con eje *X* arbitrario. Procedimientos - Recuento de resultados en el experimento - Representación del eje *X*: en el cual se representan los posibles sucesos. No está naturalmente ordenado. - Representación en el eje *Y*: frecuencias absolutas. - Representación de las barras con una altura proporcional a las frecuencias (convenio del gráfico).

En la figura 4.7 se muestra un ejemplo de un gráfico incorrecto dentro de este nivel de complejidad (nivel C1). Como ya hemos dicho, en el proyecto los alumnos analizaron tres pares de variables (número de caras, número de rachas y longitud de la racha más larga en las secuencias real y simulada), y en este caso se ve como el alumno realiza un gráfico para representar en él el valor de la variable que obtuvo para las tres variables referentes a la secuencia simulada. El gráfico elaborado es incorrecto, porque en el gráfico de sectores lo que se debe representar es las frecuencias de diferentes valores de una misma variable, es decir, una distribución de datos y no valores diferentes de tres variables. Por tanto, asumimos que el alumno tiene un conflicto con la idea de distribución, que no asocia a una única variable. Además, estas variables no son comparables entre sí, y de este gráfico el alumno no puede sacar ninguna información que le permita resolver el problema planteado. El alumno utiliza la hoja Excel para realizar su gráfico por lo que no está claro que sepa los convenios de construcción del gráfico de sectores, por ejemplo no está claro que sepa que la amplitud de cada uno de los sectores debe ser proporcional a la frecuencia. Sin embargo, al menos implícitamente está usando las ideas de círculo y sector circular.

Figura 4.7. Análisis semiótico de un gráfico incorrecto de la categoría C1

Representación de resultados en la secuencia del estudiante	*Configuración cognitiva*
 Lenguaje - Gráfico: círculo, líneas, sectores circulares. - Verbal: título del gráfico y rótulos para nombrar las distintas variables representadas. - Uso de diferentes colores para diferenciar los sectores del círculo. - Título del gráfico incompleto. - No queda claro el estadístico representado para cada variable; parece deducirse que el valor numérico de la variable en el experimento particular.	Conceptos y proposiciones - Experimento aleatorio: lanzar una moneda - Sucesos o resultados del experimento, espacio muestral. - Variables aleatorias: – A: Variable con dos valores: ñCaraò y ñCruzò (Bernoulli). – B: Número de rachas en 20 lanzamientos (valores 0-20). – C. Longitud de la racha mayor en 20 lanzamientos (valores 0-20). - Variable estadística: resultados de repetir 20 veces el experimento y contar el número de caras. - No considera las variables estadísticas correspondientes a las variables aleatorias B y C puesto que sólo considera su experimento. - Frecuencias absolutas de caras y cruces. - Valores numéricos de las tres variables en estudio en su experimento particular. - Conflicto respecto a la idea de distribución, no asocia la distribución a una variable, pues representa los valores obtenidos en su experimento en variables diferentes en el mismo gráfico. - Círculo, sector circular, amplitud. Procedimientos - Recuento de los resultados del experimento. - Introducción de los datos del experimento en Excel para realizar el gráfico. - Selección del gráfico de sectores en Excel. - Introducción de datos dentro de la función Excel que proporciona el gráfico de sectores, incluyendo el título y etiquetas. - Diferenciación de los valores de las distintas variables por medio de la utilización de distintos colores.

C2. Representa los valores individuales de la variable. Son los estudiantes que no llegan a agrupar los valores similares del número de caras obtenidos en las secuencias reales o simuladas del total de alumnos en la clase. En lugar de ello, representan el valor (o valores) obtenidos para cada alumno dentro del gráfico. Se trata de una representación de los datos en el orden en que han sido obtenidos, pero no se llega a la idea de frecuencia asociada a cada valor o de distribución de frecuencias de la variable estadística.

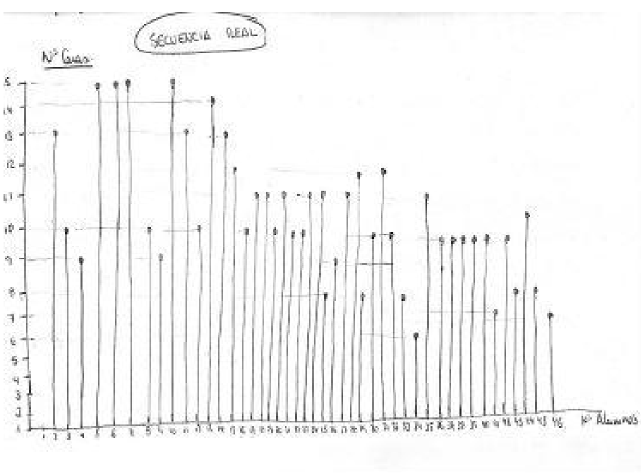
Hacemos notar que el orden de presentación de los datos en el eje *X* es artificial, pues sólo indica el orden arbitrario en que se recogieron los datos en la clase. Dentro de esta categoría hemos obtenido diagramas de barras horizontales y verticales; gráficos de líneas de una o las dos variables, que, aunque no permiten resolver el problema de comparación, al menos muestran la variabilidad de los datos en los diferentes estudiantes.

En las figuras 4.8 y 4.9 se muestran dos ejemplos de gráficos realizados por los futuros profesores, correspondientes a la categoría C2. En el primero de los casos el estudiante realiza diagramas de barras separados para cada secuencia, permitiendo este tipo de gráficos mostrar la variabilidad de los datos obtenidos por los alumnos, aunque dificultando mucho la comparación de las secuencias real y simulada para la obtención de una conclusión final en el proyecto, pues no se observa con claridad la moda ni las frecuencias de los diferentes valores. Sin embargo, usa implícitamente las ideas de experimento aleatorio, variable aleatoria y sus valores, así como la variabilidad de la variable.

El alumno usa la idea de proporcionalidad, pues la altura de cada barra corresponde al valor de la variable obtenido por cada alumno. El valor de la variable se ha representado en el eje *Y*, que usualmente corresponde a las frecuencias. En el eje *X* se representan los números naturales ordenados en su orden natural. Por tanto establece una correspondencia entre las divisiones en el eje que son los números naturales y cada uno de los alumnos, aunque dicha correspondencia es arbitraria, así en el eje *X* se representan los alumnos de la clase ordenados según su colocación en el aula, es por tanto un orden arbitrario que no corresponde al orden de los números naturales.

También se visualiza la idea de variable estadística (diferentes valores de la variable aleatoria para cada estudiante), pues se representa cada unidad estadística (alumno) con su valor de la variable.

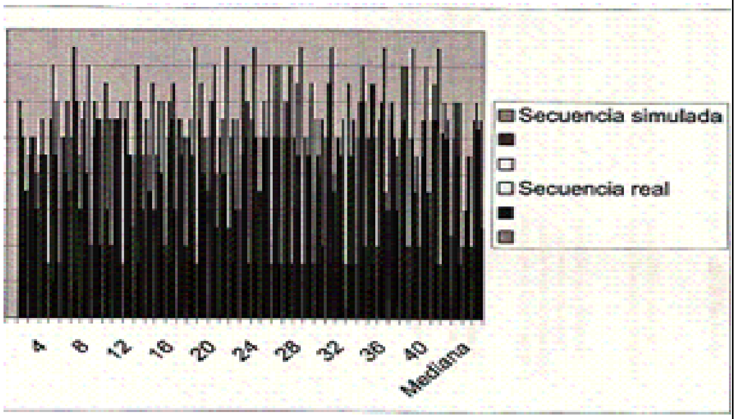
Figura 4.8. Análisis semiótico de un gráfico correcto del tipo C2

Gráfico de barras separados de cada secuencia	*Configuración cognitiva*
 Procedimientos - Representa en el eje *X*: Número de orden de los alumnos en la clase, según su colocación, no corresponde a una propiedad del conjunto de alumnos. - Establece una correspondencia entre las divisiones en el eje que son los números naturales y cada uno de los alumnos. Dicha correspondencia es arbitraria. - Representación del eje *Y*: número de caras obtenido por cada alumno, es decir representa el valor de la variable. Hay divisiones, en el eje que representan los resultados individuales obtenidos por cada uno de los alumnos.	Conceptos y proposiciones - Experimento aleatorio: lanzamiento de una moneda 20 veces. - Variables aleatorias: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 por un estudiante genérico. - Variables estadísticas: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 obtenidas por cada uno de los *m* estudiantes de la clase. - Valor de la variable aleatoria para cada alumno. - Proporcionalidad (altura de la barra con valor de la variable). - Números naturales. - Orden de los números naturales. - Segmento. - Perpendicularidad y paralelismo. - Variabilidad de los datos. Lenguaje - Grafico: líneas, puntos, marcas. - Verbal: título del gráfico y etiquetas de los ejes. - Numérico, en las escalas de los ejes. - Rótulo para el título del gráfico confuso al no hacer referencia a la variable ñNº de carasò. - Incluye rótulo en los dos ejes.

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.

En la figura 4.9 el estudiante representa un gráfico de barras adosado para ambas secuencias. En este ejemplo el gran número de barras en el gráfico dificulta la lectura y comprensión del gráfico e incluso impide la visualización de la variabilidad de los datos por lo cual consideramos que el gráfico es incorrecto. Tanto para la secuencia real como la simulada aparecen representados los valores individuales de las variables aleatorias para los diferentes estudiantes y estos se presentan respondiendo a un orden artificial, no corresponde a una propiedad del conjunto de los alumnos sino al orden en que fueron obtenidos los datos en la clase. Puesto que el gráfico está hecho con la hoja Excel no es claro que el estudiante haya usado explícitamente la idea de proporcionalidad.

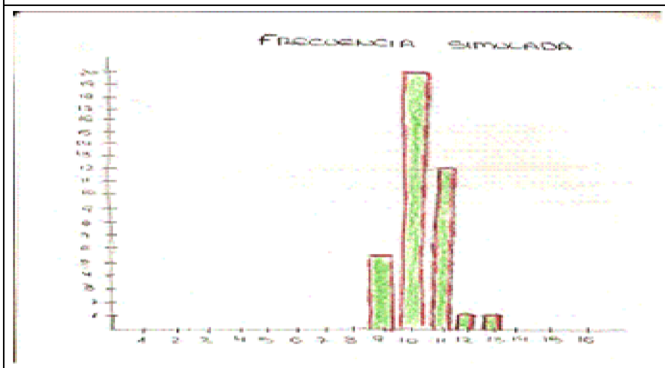
Figura 4.9. Análisis semiótico de un gráfico incorrecto del tipo C2

Gráfico de barras adosado de las dos secuencias	*Configuración cognitiva*
 Lenguaje - Grafico: líneas, marcas, cuadrados (para representar las series). - Verbal: rótulo de uno de los ejes, rótulos para las series. - Numérico: en la escala del eje *Y*. - No incluye el título del gráfico. - No hay rótulos en el eje *Y*. - El rótulo del eje *X* no corresponde a lo representado (número de alumno). Procedimientos - Representa en el Eje *X*: Número de orden de los alumnos en la clase, según su colocación. El orden es artificial, no corresponde a una propiedad del conjunto de alumnos. - Representa en el eje *Y*: el valor de la variable. - Ha tenido que introducir los datos en la Hoja Excel, procedimiento de grabación de datos. - Representa varias series en el Excel, dos de ellas para diferenciar los valores de la secuencia real y la simulada, representadas a la derecha del gráfico con cuadraditos de distintos colores.	Conceptos y proposiciones - Experimento aleatorio: lanzar una moneda 20 veces. - Variables aleatorias: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 por un estudiante genérico. - Variables estadísticas: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 obtenidas por cada uno de los *m* estudiantes de la clase. - Valores de la variable para cada alumno. - Hay un *conflicto semiótico* sobre la idea de mediana pues confunde mediana con número del estudiante. - Proporcionalidad de la altura de la barra con valor de la variable. Aunque al ser un gráfico realizado con el Excel no se puede saber si el alumno utiliza este concepto explícitamente. - Números naturales y orden de los números naturales.

En este ejemplo (figura 4.9), las etiquetas de los ejes son claramente deficientes, pues el eje Y no presenta los números naturales ordenados y los que aparecen en el eje X no corresponden a lo representado (estudiantes de la clase). Hay un conflicto semiótico pues el estudiante confunde al menos a nivel verbal la *mediana* con unidad estadística (número del estudiante). En este ejemplo, así como en el presentado en la figura 4.7 se muestra como el uso del ordenador ha dificultado la construcción del gráfico por parte del alumno. En este ejemplo, parece que el alumno quiere representar los valores individuales de las variables en un mismo gráfico, pero que acepta las opciones que vienen por defecto en la hoja Excel, así, según la categorización dada por Ben-Zvi y Friedlander (1997), este gráfico correspondería a un *Uso acrítico*, por parte del estudiante, del software para la realización del gráfico.

C3. *Produce gráficos separados para cada distribución.* El alumno forma una tabla de frecuencias de cada una de las dos variables y a partir de ella produce un gráfico o bien representa directamente un gráfico de cada uno de los valores diferentes de la variable (para las dos variables) con sus frecuencias. Esto supone que el alumno pasa del conjunto de datos a la variable aleatoria ñnúmero de caras en 20 lanzamientosò y a su correspondiente variable estadística con su distribución de frecuencias en la muestra de *m* estudiantes. El usar dos gráficos separados dificulta a veces la comparación de las variables, sobre todo en caso de no usar la misma escala de representación en los dos gráficos o usar distintos tipos de gráficos para cada una de las secuencias.

Figura 4.10. Análisis semiótico de un gráfico correcto del tipo C3

- *Gráficos de barras correctos*	*Configuración cognitiva*
 Lenguaje - Grafico: rectángulos, líneas, marcas. - Verbal: título del gráfico. - Numérico: etiquetas de las escalas, que representan los números naturales ordenados. - No hay rótulos en los ejes. - El título del gráfico es confuso, pues lo que se representa es la distribución de frecuencias del número de caras en la secuencia simulada. Procedimientos - Recuento de datos, cálculo de frecuencias. - Representación de los ejes: en el eje *X* representación de los valores de la variable (número de caras) y en el eje *Y* representación de las frecuencias asociadas a cada valor de la variable. - Representación de escalas en los ejes. - Representación de frecuencias mediante barras.	Conceptos y proposiciones - Experimento aleatorio: lanzar 20 veces una moneda. - Variables aleatorias: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 por un estudiante genérico. - Variables estadísticas: muestra de tamaño m de los resultados del número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 obtenidas por cada uno de los m estudiantes de la clase. - Frecuencias de cada posible valor de la variable. - Proporcionalidad (altura de la barra con frecuencia). - Números naturales y orden de los naturales. - Paralelismo y perpendicularidad. - Distribución de la variable (conjunto de valores con su frecuencia). - Moda (valor más frecuente). - Mínimo, máximo y rango.

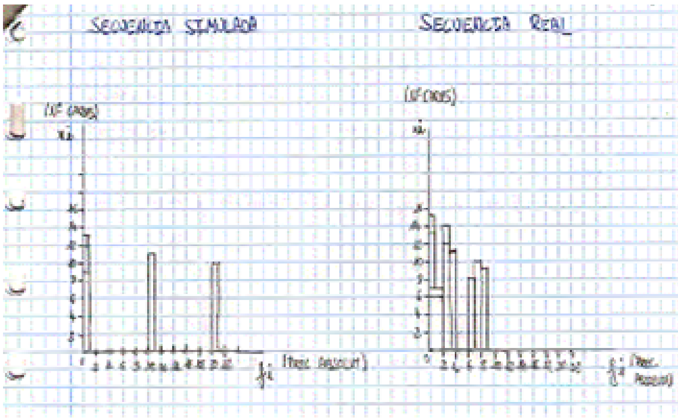
Entre los gráficos correctos dentro de esta categoría hemos obtenido gráficos de barras horizontales y verticales y polígonos de frecuencias. En las figuras 4.10 y 4.11 se muestran, respectivamente, un gráfico correcto y otro incorrecto pertenecientes al nivel de complejidad semiótico C3. En ambos casos los alumnos llegan a la distribución de frecuencias de las variables estadísticas, pasando de los datos a la variable y su distribución.

En la figura 4.10 el estudiante ha sido capaz de agrupar los datos obtenidos en el experimento realizado por la clase según el valor de la variable, llegando así al concepto de variable estadística, frecuencia asociada a cada uno de los posibles valores de ésta y a su distribución. En este caso el gráfico utilizado es un gráfico de barras, en el que cada barra representa la frecuencia asociada a cada valor de la variable estadística y en cuya representación el alumno utiliza la idea de proporcionalidad. Este gráfico permite visualizar fácilmente la moda, el valor máximo y mínimo de la variable y por tanto el rango, por lo que si en el gráfico análogo para la secuencia real se utilizasen las mismas escalas, sería posible extraer interpretaciones y conclusiones de la comparación de dichos resúmenes estadísticos en ambos gráficos. Tanto en el eje X como en el Y el alumno representa los números naturales ordenados, sin poner rótulos a los ejes. Además el título del gráfico puede considerarse incompleto al no hacer referencia a la variable ñNº de caras en la secuencia simuladaò cuya distribución está siendo representada.

Los alumnos también producen gráficos incorrectos, como histogramas con intervalos incorrectamente representados, gráficos de barras con ejes intercambiados (confusión de la variable dependiente e independiente en la distribución de frecuencias), representación de las frecuencias adosadas a los valores de la variable (no se comprende el significado de la frecuencia) y gráficos de sectores, que no permiten mostrar la variabilidad de los datos. En el ejemplo de la figura 4.11 se observa un gráfico incorrecto dentro de esta categoría. Se trata de un gráfico de barras en el que no se han respetado los convenios de construcción de este tipo de gráficos ya que, han sido intercambiadas las variables independiente (número de caras) y dependiente (frecuencias absolutas asociadas al número de caras) de la distribución de frecuencias en los ejes.

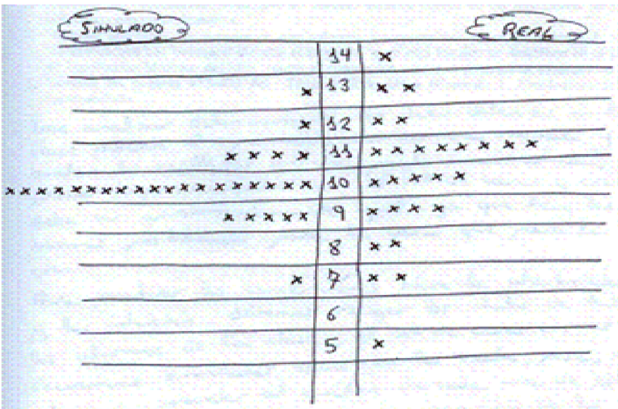
Este conflicto ya fue detectado por Ruiz (2006) en un estudio de comprensión de la variable aleatoria y lleva consigo una mayor dificultad en la lectura e interpretación del gráfico. En el ejemplo mostrado no se perciben todos los valores diferentes de la variable, pues si hay dos valores con la misma frecuencia, se representan sobre la misma barra. Además, la moda en este caso es más complicada de detectar visualmente. Esto conlleva una mayor dificultad en la comparación de los gráficos de la secuencia simulada y la secuencia real y por lo tanto en la interpretación y extracción de conclusiones para dar respuesta a la pregunta planteada en el proyecto. A pesar de estos problemas el alumno es capaz de agrupar los datos brutos y llegar al concepto de variable estadística y frecuencias y utilizar la idea de proporcionalidad en la construcción del gráfico.

Figura 4.11. Análisis semiótico de un gráfico incorrecto del tipo C3

Gráfico de barras con ejes cambiados	*Configuración cognitiva*
 Lenguaje - Grafico: rectángulos, líneas, marcas. - Verbal: rótulos de los ejes y títulos de ambos gráficos. - Numérico: Escalas en los ejes que representan los números naturales ordenados; incluye etiquetas de los valores. - Simbólico: utiliza el símbolo x_i para referirse a los posibles valores de la variable estadística y f_i para referirse a las frecuencias absolutas de los distintos valores que toma la variable. - El título del eje es incompleto, pues lo que se representa es la distribución del número de caras en ambas secuencias. Procedimientos - Recuento de datos y cálculo de frecuencias. - Representación de los ejes: en el eje *Y* representa el número de caras (valores de la variable estadística) y en el eje *X* la frecuencia asociada a cada valor de la variable. - Representación de escalas en los ejes. - Representación de valores de la variable mediante barras. - Utiliza la misma escala en ambos gráficos lo cual facilita la comparación. - No sigue el convenio en la representación de las frecuencias	Conceptos y proposiciones - Experimento aleatorio: lanzar 20 monedas. - Variables aleatorias: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 por un estudiante genérico. - Variables estadísticas: muestra de tamaño *m* de los resultados del número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 obtenidas por cada uno de los *m* estudiantes de la clase. - Frecuencias de cada posible valor de la variable. - Distribución de frecuencias. - Paralelismo y perpendicularidad. - Proporcionalidad (altura de la barra con valor de la variable). - Números naturales y orden de los naturales. - *Conflicto* al confundir frecuencia y valor de la variable. - *Conflicto* al confundir la variable dependiente e independiente en la distribución de frecuencias.

C4. Produce un gráfico conjunto de las dos distribuciones. El alumno ha llegado a formar las distribuciones de las dos variables y las representa conjuntamente en el mismo gráfico, lo cual facilitará su comparación. El gráfico tiene también mayor complejidad al representar conjuntamente dos variables estadísticas. Hemos encontrado las siguientes variaciones de gráficos correctos: diagramas de barras adosados, representación de algún estadístico comparable en las dos distribuciones (por ejemplo las medias), gráficos de barras contrapuestos de estadísticos comparables en las dos muestras (por ejemplo, media, mediana y moda), gráficos de líneas superpuestos y gráficos de puntos de las dos distribuciones. Entre los gráficos incorrectos, los alumnos representan algún estadístico (por ejemplo las medias) en variables no directamente comparables, o de varios estadísticos no comparables en las dos variables (por ejemplo, media y rango) o diagrama de sectores del total de caras en las dos distribuciones.

Figura 4.12. Análisis semiótico de un gráfico correcto del tipo C4

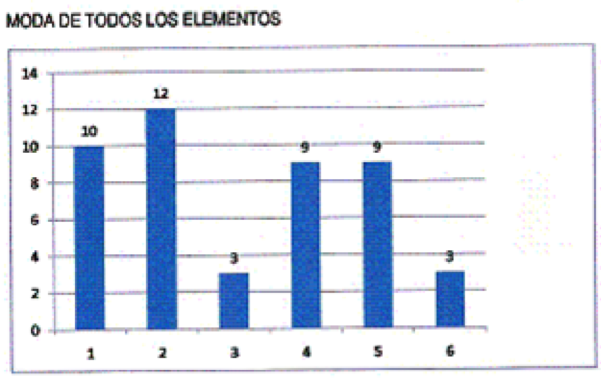
<table>
<tr><th>Gráfico de puntos doble</th><th>Configuración cognitiva</th></tr>
<tr><td>


Lenguaje
- Gráfico: cruces, líneas.
- Verbal: rótulos para cada una de las distribuciones, incompletos ya que sólo dan información de la secuencia de la que se trata y lo que representan son las distribuciones del número de caras.
- Numérico: etiquetas para la escala del eje Y que son los números naturales ordenados.
- En el eje Y debemos interpretar que representa los valores de las variables, ya que no tiene rótulo.
Procedimientos
- Clasificación de cada valor de la variable para colocar un punto en el lugar del gráfico que le corresponde (no hay recuento explícito ni cálculo explícito de frecuencias)
- Representación del eje Y y su escala.
- Representación de los valores de las variables mediante puntos.
</td><td>
Conceptos y proposiciones
- Experimento aleatorio: lanzamiento de una moneda equilibrada 20 veces.
- Variables aleatorias: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 por un estudiante genérico.
- Variables estadísticas: muestra de tamaño m de los resultados del número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 obtenidas por cada uno de los m estudiantes de la clase.
- Frecuencias de cada posible valor de la variable (implícito, dado por el número de puntos)
- Distribución de frecuencias.
- Sucesos, espacio muestral.
- Números naturales y orden de los números naturales.
- Moda (valor más frecuente).
- Valores máximo y mínimo de las variables y sus rangos.
- Mediana.
- Conflicto: cada punto representa el nº de caras obtenido por un determinado alumno, ya sea en la secuencia real o simulada, pero en la secuencia simulada aparecen más puntos cosa que no es posible.
</td></tr>
</table>

En la figura 4.12 se muestra un ejemplo correcto de un gráfico de esta categoría realizado por un alumno: un gráfico de puntos en el que se muestra la distribución de las dos variables en estudio. En este gráfico al igual que ocurría con el mostrado en la figura 4.10 que correspondía a un gráfico correcto de la categoría C3, el alumno es capaz de agrupar los datos iguales llegando así al concepto de variable estadística, frecuencias asociadas a los valores de dichas variables y sus distribuciones de frecuencias. Este gráfico supone una mayor complejidad que el de nivel C3 pues maneja dos variables estadísticas en el mismo gráfico. También permite una más fácil interpretación de los contenidos estadísticos subyacentes en el, ya que posibilita visualizar la moda, los valores máximos y mínimos de las variables, los rangos e incluso permite, por un simple mecanismo de conteo de puntos, calcular cual es la mediana en ambas distribuciones. Este gráfico tiene la ventaja de facilitar mucho la comparación de ambas distribuciones. El alumno hace un pequeño error en el gráfico pues no ha colocado el mismo número de puntos en cada parte del gráfico, mientras que el número de datos de las dos variables estadísticas es el mismo. Pensamos que este es un error accidental por lo que consideramos el gráfico correcto.

Figura 4. 13 Análisis semiótico de un gráfico incorrecto del tipo C4

- *Gráfico de barras de las modas de tres pares de variables*	*Configuración cognitiva*


 	
Lenguaje
- Gráfico: rectángulos, líneas, marcas
- Verbal: título del gráfico, aunque confuso, ya que se representa la moda de los tres pares de variables estudiadas en el proyecto.
- Numérico: etiquetas en ambos ejes que son los números naturales ordenados.; valores de las modas encima de cada barra
- Las etiquetas del eje X son confusas pues no indica que cada número natural se corresponde con cada una de las variables estadísticas.
Procedimientos
- Representación de los ejes: En el eje Y se representa el valor de la moda de cada una de las variables, en el eje X cada número representa una de las 6 variables del proyecto.
- Representación de la moda de los tres pares de variable mediante barras y mediante un valor numérico asociado a cada barra.
	Conceptos y proposiciones
	- Experimento aleatorio: lanzamiento de una moneda equilibrada 20 veces.
	- Variables aleatorias: número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 por un estudiante genérico.
	- Variables estadísticas: muestra de tamaño *m* de los resultados del número de caras al lanzar 20 veces una moneda equilibrada y número de caras en la secuencia inventada de longitud 20 obtenidas por cada uno de los *m* estudiantes de la clase.
	- Variables aleatorias:
	  - Número de rachas en 20 lanzamientos de una moneda (real o simulada).
	  - Longitud de la racha más larga en 20 lanzamientos de una moneda (real o simulada).
	- Variables estadísticas asociadas a las anteriores variables aleatorias
	- Moda.
	- Números naturales y orden de los números naturales.
	- Al haber realizado el gráfico con Excel, no está claro que el estudiante haya usado explícitamente los conceptos de paralelismo, perpendicularidad y proporcionalidad.

En la figura 4.13 se muestra un ejemplo de gráfico elaborado por un alumno perteneciente a la categoría C4, que es la de mayor complejidad semiótica. Dicho gráfico es incorrecto, ya que en él el alumno representó la moda de los tres pares de variables puestas en juego en el proyecto que elaboraron (número de caras, número de rachas y longitud de la racha más larga en ambas secuencias, real y simulada) y estas no son todas comparables entre sí. Pero, a pesar de ser incorrecto, el alumno en dicho ejemplo ha tenido que pasar también de los datos brutos a las variables estadísticas del estudio y su distribución, y a calcular resúmenes estadísticos, en este caso la moda. Otro error que se puede observar es que no pone rótulos a ninguno de los ejes, el eje *Y* debido al título del gráfico se puede interpretar fácilmente que representa la moda de determinadas variables estadísticas, sin embargo en el eje *X*, se ha de suponer que cada uno de los números representa una de las variables estadísticas anteriormente citadas, además el orden en este eje es artificial.

Entre cada uno de los niveles semióticos antes definidos, se evidencia un salto cualitativo, a pesar de que del análisis de los resultados de los alumnos, se puede observar una gran variedad de gráficos y configuraciones de objetos diferenciados, e incluso gráficos correctos e incorrectos.

4.4.2. CORRECCIÓN DE LOS GRÁFICOS.

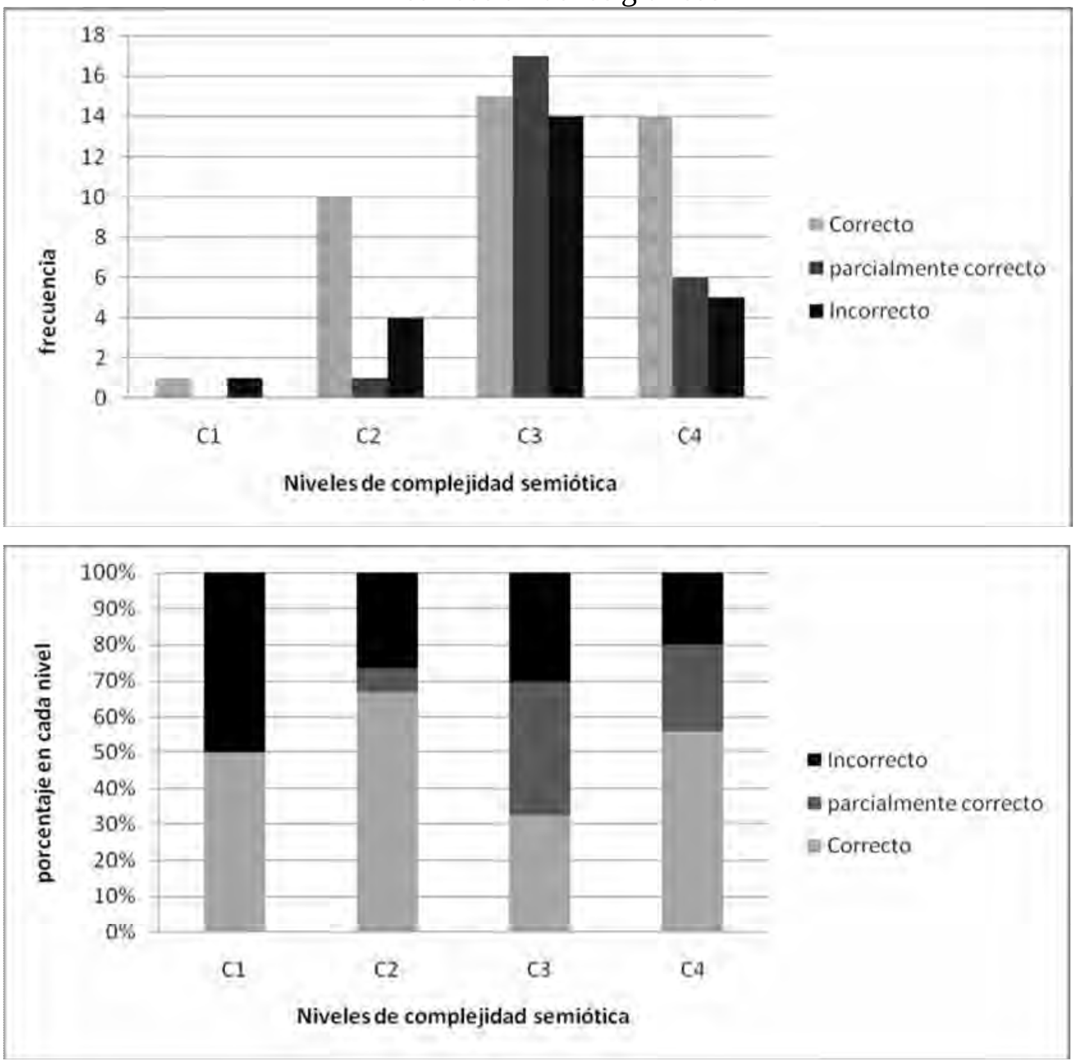
Un segundo punto analizado es la corrección de los gráficos (Arteaga, Ortiz y Batanero, 2008). Del total de 93 alumnos 88 (94,6%) producen algún tipo de gráfico para analizar sus datos, incluso cuando las instrucciones de la tarea no los requerían. Este alto porcentaje indica la necesidad sentida de los estudiantes de producir un gráfico y llegar, mediante un proceso de transnumeración, a un conocimiento no disponible en los datos brutos.

Tabla 4.4. Distribución 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.	
C1. Representa sólo sus datos	1		1	2
C2. Representa resultados individuales	10	1	4	15
C3. Gráficos separados para cada muestra	15	17	14	46
C4. Gráficos conjuntos	14	6	5	25
Total	40	24	24	88

En la Tabla 4.4 se presenta la clasificación de los alumnos según el nivel de los gráficos que elaboran y la corrección de los mismos. En la figura 4.14 se representan estos datos, primeramente mediante un gráfico de barras adosado de frecuencias absolutas y seguidamente, mediante un gráfico apilado de barras en porcentajes dentro de cada categoría.

Figura 4.14. Distribución de estudiantes, según nivel de complejidad semiótica y corrección de los gráficos



En dicha tabla y gráficos se observa que la mayoría de los estudiantes que representan los datos gráficamente (52,2%), producen gráficos separados para cada variable (nivel C3), generalmente correctos o parcialmente correctos. Consideramos que el gráfico es parcialmente correcto si no teniendo errores esenciales, el alumno no usa la misma escala en los dos gráficos o bien no usa el mismo gráfico en las dos muestras, dificultándose por ello la comparación. Otros casos considerados parcialmente correctos es cuando no centran el intervalo en los histogramas, no hay coincidencia de los valores representados con la escala utilizada o no incluyen etiquetas. Catorce alumnos en este nivel producen gráficos sin sentido al representar en el mismo gráfico los promedios del número de caras y de otras variables utilizadas en la realización del proyecto, las cuales no han son objeto de estudio en este trabajo (número de rachas y longitud de la racha más larga) , representan 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 dependiente error ya detectado en Ruiz (2006) en relación al concepto de variable aleatoria.

El 28,4% trabajan al nivel C4, 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. Este nivel tiene un alto porcentaje de gráficos correctos (56%) y si tenemos en cuenta también los gráficos parcialmente correctos se tiene que un total del 80% de los gráficos pertenecientes a esta categoría, son al menos, parcialmente correctos. Los alumnos que llegan a construir gráficos del nivel C4, a pesar de ser el nivel de mayor complejidad semiótica, consiguen, en su mayoría, realizar el gráfico de manera que sea al menos correcto parcialmente.

Son pocos los estudiantes que analizan sólo sus propios datos (nivel C1) y sólo 17% estudian los valores obtenidos por cada estudiante caso a caso sin llegar a formar la distribución, por lo que dicho concepto parece ser adquirido y utilizado por los estudiantes para resolver la tarea propuesta. En este caso (nivel C2), la mayoría de los gráficos producidos son correctos, precisamente el mayor porcentaje de gráficos construidos correctamente se da en el nivel C2, siendo un 62,5% del total de los gráficos de esta categoría correctos. Al ser gráficos muy simples, el alumno no produce errores en la representación de los datos; sin embargo, puesto que el alumno no ha llegado a la distribución de frecuencias de la variable, estos gráficos no permiten resolver el problema propuesto.

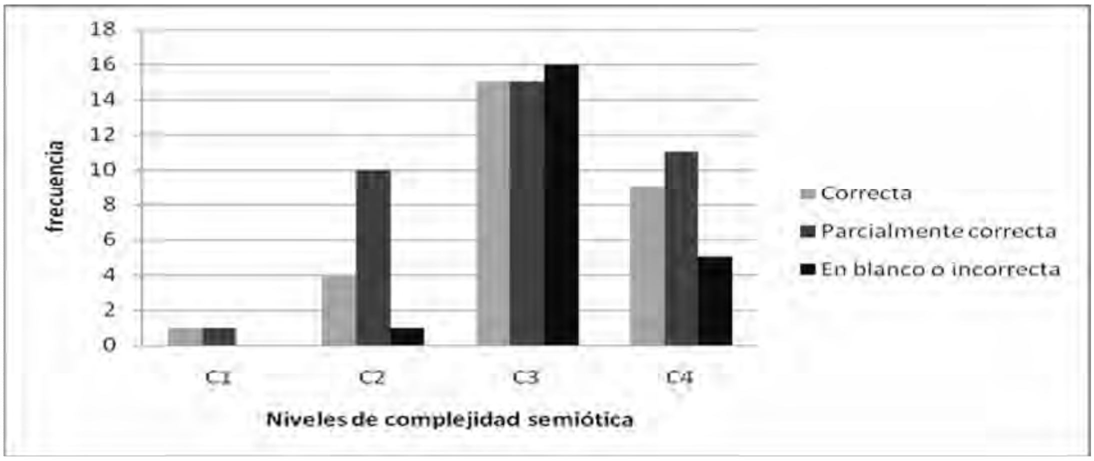
4.4.3. INTERPRETACIÓN Y NIVELES DE LECTURA

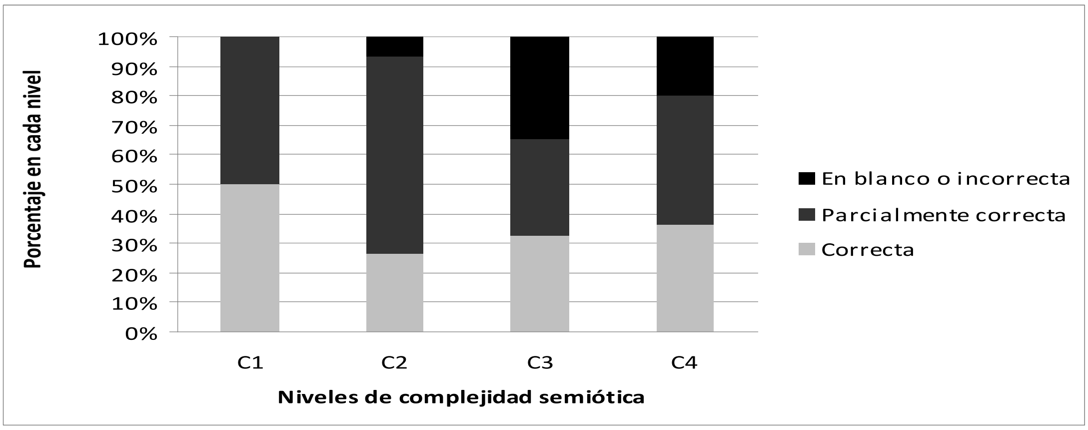
Hemos también analizado la capacidad de lectura de los gráficos producidos por parte de los estudiantes, para ver si se alcanzan los niveles superiores de comprensión de gráficos definidos por los autores que hemos citado en los antecedentes. En la Tabla 4.5 y Figura 4.15 hemos clasificado los estudiantes, según nivel del gráfico e interpretación que hacen del mismo.

Tabla 4.5 Distribución de estudiantes, según nivel de complejidad semiótica e interpretación de los gráficos

Categorías de gráfico construido	Interpretación del gráfico			Total en el nivel
	Corr.	Parcial	Blanco o Inc.	
C1. Representa sólo sus datos	1	1		2
C2. Representa resultados individuales	4	10	1	15
C3. Gráficos separados para cada muestra	15	15	16	46
C4. Gráficos conjuntos	9	11	5	25
Total	29	37	22	88

Figura 4.15. Clasificación de estudiantes, según nivel de complejidad semiótica e Interpretación de los gráficos





Interpretaciones incorrectas o no interpretación

Elaborados los gráficos, algunos estudiantes se limitan a presentarlos y no los comentan. En otros casos se hace una interpretación incorrecta, por ejemplo el alumno que elabora la gráfica 4.11, hace la siguiente interpretación:

ñEn la secuencia real las frecuencias son más consecutivas, no se repite la misma cara tantas veces y se repiten menos veces y distintos números, por ello las barras están más juntasò (Alumno CB).

La confusión entre valores de la variable y frecuencia, hace que compare las frecuencias (valores representados en el eje X). Puesto que en las secuencias reales hay más valores diferentes de la variable, las frecuencias de cada valor son menores y también hay más frecuencias diferentes. Al representar los valores que aparecen en secuencia real (como el 13) pero no en la simulada con frecuencia nula para las simuladas y también haber obtenido una frecuencia de 20 alumnos con 10 caras, los valores de las frecuencias en este gráfico son más diferentes que en la secuencia real. Es lo que quiere expresar al decir que en un gráfico ñlas frecuencias son más consecutivasò. La interpretación tiene varios errores, además de los señalados, pues confunde ñconsecutivoò con ñdistanteò o ñseparadoò. Por otro lado, para que las variables fueran diferentes debieran ser diferentes sus valores, o al menos las frecuencias para el mismo valor, no las frecuencias aisladas.

Otro ejemplo de interpretación errónea es el siguiente, pues las medias en ambas distribuciones (y en el gráfico concreto que produce el estudiante) son cercanas al valor 10 y no cercanas a cero como argumenta el estudiante. Vemos que también el estudiante está atribuyendo la idea de representatividad a la desviación típica, cuando esta es una propiedad de las medidas de posición central y no de las de dispersión.

ñAnalizando la variable ñnúmero de carasò el rango de la secuencia simulada permite apreciar más la uniformidad de la media que en la secuencia real donde el rango es 9; a pesar de esto en ambas desviaciones típicas se ve una gran representatividad ya que ambos valores de las medias están muy próximos a ceroò (Alumno ER).

En este otro ejemplo, el alumno hace una lectura incorrecta del mínimo (que es 9) en la gráfica producida:

ñEn la experiencia de la secuencia simulada, el número de caras que cada uno de la clase pensábamos estaban a partir de 8 carasò (Alumno TF)

Destacamos, en primer lugar que una cuarta parte de los estudiantes se limita a producir el gráfico sin llegar a interpretarlo o bien da una interpretación errónea, pues, aunque leen elementos aislados del gráfico, no llegan a la extracción de tendencias o análisis de la estructura (Bertín, 1967) en su nivel de lectura de gráficos. Estos errores o esta falta de interpretación se produce sobre todo en los gráficos de nivel C3 y C4, pues la menor complejidad semiótica de los gráficos de nivel C1 y C2 hace que estos sean más sencillos de interpretar para los estudiantes.

Interpretaciones parcialmente correctas

Un 42% de los estudiantes hacen una interpretación parcial de los gráficos, analizando tan sólo los promedios sin tener en cuenta la variabilidad o bien al contrario, comparando sólo los rangos de variación. Estos alumnos, según la clasificación de niveles de lectura de los gráficos dada por Bertin (1976), ya habrían superado el nivel inicial de ñextracción de datosò encontrándose en el nivel inmediatamente superior de ñextracción de tendenciasò, siendo capaces de percibir características de la distribución a partir de la lectura de los gráficos, como puede ser la mayor variabilidad de la secuencia real o detectar visualmente el valor más frecuente o moda.

En el siguiente ejemplo la interpretación es parcialmente correcta porque la alumna compara sólo la dispersión y efectivamente hay más variabilidad en las secuencias reales. La alumna había hecho un gráfico de nivel C2, representando los datos alumno a alumno sin formar la distribución. Pero como usa un gráfico de líneas este permite ver la mayor variabilidad en una de las dos series, porque en este caso las oscilaciones del gráfico son más amplias. Lo consideramos parcialmente correcto porque no compara los promedios.

> ñEn cuanto al número de caras reales que ha obtenido el grupo es mucho más variable son más homogéneos que en la simulación. Para poder realizar esta afirmación hemos usado el rangoò (Alumno JI).

Un argumento muy similar lo da el alumno LC, aunque es más preciso que el anterior al indicar la proporción de las diferencias en dispersión (seis veces más) y hacer referencia a la varianza y desviación típica que también calculó para sus datos.

> ñEn primer lugar comparé los datos del número de caras en las dos secuencias donde destacaré las diferencias que se aprecian principalmente entre el rango, la varianza y consecuentemente la desviación típica de hasta seis veces más en la secuencia simulada respecto a la realò (Alumno LC).

El alumno AG por el contrario usa un vocabulario más impreciso; sin embargo se ven puntos de interpretación correcta del gráfico (en el diagrama de líneas al ascender la línea aumenta la frecuencia; detección de la moda). En este caso se comparan las modas y porcentajes de valores pero no la dispersión.

> ñEl sistema de representación de datos utilizado es el diagrama de líneas ya que se pueden ver las diferencias entre la secuencia simulada y la real; si la línea asciende, la frecuencia aumenta,.. El punto más alto del diagrama corresponde al valor de la variable que tiene mayor frecuencia, como ocurreò (pone un ejemplo) (Alumno AG).

La mayor proporción absoluta de los que hacen esta interpretación parcial se encuentra entre los estudiantes que producen un gráfico de nivel C3 igualando en este nivel al número de interpretaciones correctas. En proporción relativa, dentro de cada nivel, los que producen mayor número de interpretaciones parcialmente correctas son los de nivel C2 y 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, que los estudiantes en este nivel no han sabido todavía relacionar. Por ello en la interpretación de la gráfica, bien no interpretan los promedios, bien no interpretan la dispersión.

Hay también una alta proporción de interpretaciones parcialmente correctas dentro de los que llegan a producir gráficos de nivel de complejidad C4; en algunos casos los gráficos son producidos con la ayuda de Excel, por lo que el alumno aunque llega al mayor nivel a la hora de representar los datos, no comprende bien el gráfico producido y no llega a interpretarlo en forma completa.

Interpretaciones correctas

Una tercera parte de los estudiantes conjuga estas dos ideas detectando las tendencias y analizando la estructura de los datos, llegando al nivel de ñanálisis de la estructuraò definido por Bertin (1976), que coincide con el nivel de lectura de los gráficos ñleer más allá de los datosò dado por Curcio (1989), aumentando la proporción de los que lo hacen en los dos niveles superiores.

En el siguiente caso, aunque no muy bien expresado, el alumno compara tanto la moda (tendencia central) como la dispersión (rango) dando una interpretación correcta de cada una de estas medidas.

> ñEl rango de la secuencia simulada ha sido 3 y de la secuencia real 8, por tanto podríamos deducir la mayor variedad de que salga cara en la secuencia real. La moda es de 10 en la secuencia simulada y 9 en

la secuencia real, es decir, en ambos casos se han repetido casi el mismo número de caras seguidasò (Alumno NG).

Otro ejemplo (Alumno MC) hace referencia a los valores centrales y la distribución de los datos (dispersión) alrededor de los mismos:

ñAl comparar los gráficos podemos observar que el número de caras de la simulación real está mucho más repartido que en la simulación, cuyos datos están más centrados en los números 10 y 11 al igual que en la realò (Alumno MC).

Esta misma idea se expresa en forma diferente por otros estudiantes. En el ejemplo, aunque el alumno es impreciso en el lenguaje (intervalo en lugar de valor; frecuencia, en lugar de secuencia, etc.) hace referencia a la variabilidad (fluctuación), rango de variación (8-15); moda (10 es la secuencia con mayor número de veces):

ñPara comentar las diferencias, el gráfico aporta más posibilidades pues visualmente podemos percibir las diferencias. Centrándome en el gráfico, observamos que los valores de la frecuencia simulada e concentran en torno a unos mismos intervalos, 10 y 11 mientras que en la frecuencia real los valores están más disgregados aunque también vemos que se concentran en los niveles 8-15 fluctuando al alza y a la baja. De esta forma en la frecuencia real comprobamos que 10 es la secuencia con mayor número de veces, saliendo cara, tanto en la simulada (20) como en la real (14) (sigue comparando las frecuencias de los diferentes valores)... de esta forma percibimos que la horquilla de la secuencia real es mucho más amplia que la simulada ñ (Alumno FL).

En el siguiente caso hay mejor uso del lenguaje matemático haciéndose referencia explícita al rango e implícita a la moda.

ñComparando ambas secuencias observo que se asimilan mucho en el número de caras en el que todo es igual (se refiere a las modas) a excepción del rango que en uno es de 9-12 y en otro de 6-15 ñ (Alumno PF).

Sin tener en cuenta el nivel C1, en el que solamente se encuentran dos alumnos, y uno de ellos interpretó su gráfico correctamente, con lo que el porcentaje de interpretaciones correctas de los gráficos en esta categoría es del 50%, el nivel en el cual hay mayor porcentaje de interpretaciones correctas es en el nivel C4. Una vez que el alumno es capaz de construir correctamente un gráfico de este nivel de complejidad, este tipo de gráficos facilita la comparación de tendencias en ambas distribuciones, e incluso las posibles diferencias existentes en promedios y en dispersión de estas. Dicho de otro modo el mayor nivel alcanzado en la construcción de gráficos parece implicar mayor nivel en la lectura de los mismos.

4.4.4. CONCLUSIONES DE LOS ESTUDIANTES

Finalmente, analizamos las conclusiones que los estudiantes obtuvieron respecto a la pregunta planteada, esto es, respecto a las intuiciones del grupo clase sobre el experimento aleatorio que llevaron a cabo. Hacemos notar que esta es la pregunta que origina el proyecto de análisis de datos, la organización en clase del experimento y la recogida y análisis de los datos. En la clase se ha montado una pequeña investigación para tratar de responder una pregunta. La fase final es proporcionar respuesta a dicha pregunta.

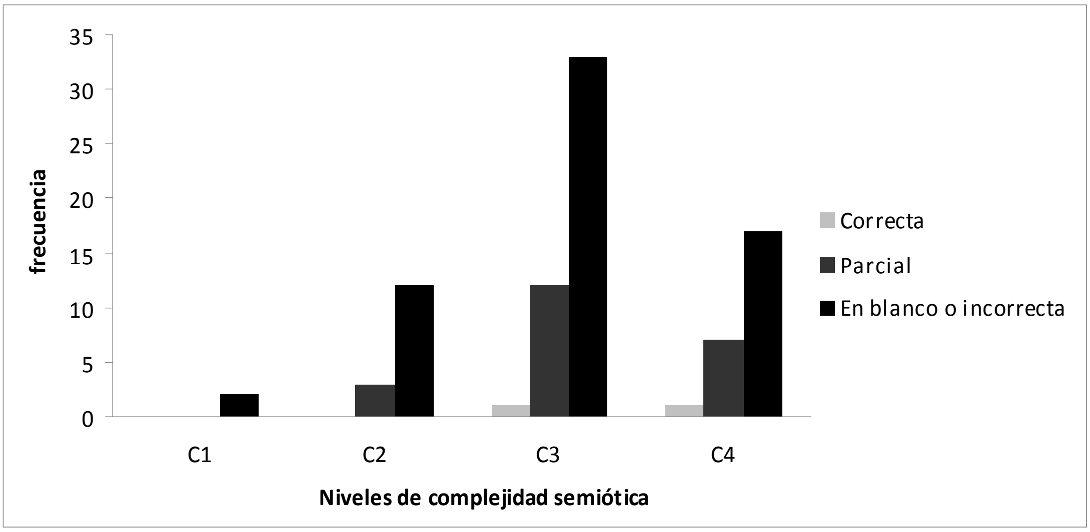
En la tabla 4.6 y figura 4. 16. presentamos las conclusiones obtenidas en relación con las intuiciones de la clase respecto a los fenómenos aleatorios. Desafortunadamente, una parte de los estudiantes se limita a presentar los gráficos y cálculos estadístico y no finaliza el proyecto, es decir, no responde a la pregunta de investigación planteada.

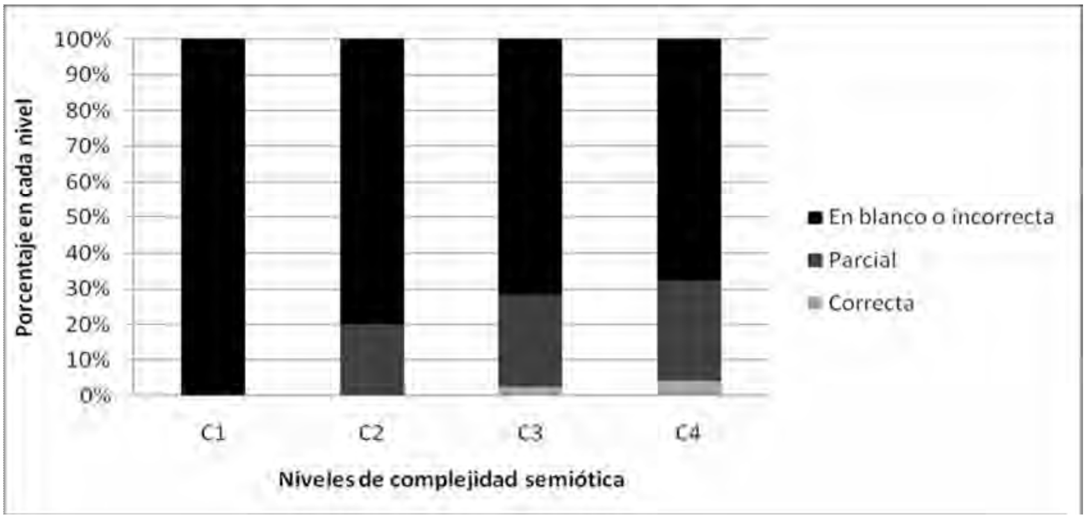
En otros casos las conclusiones son erróneas, bien por no haber entendido el propósito del proyecto o bien porque el estudiante muestra una concepción errónea de los experimentos aleatorios que interviene en su conclusión. Observamos que la obtención de la conclusión es la tarea difícil para todos los estudiantes, siendo sólo una tercera parte los que obtienen una, al menos parcial y creciendo esta proporción, con el nivel de complejidad semiótica del gráfico.

Tabla 4.6 Distribución de estudiantes, según nivel de complejidad semiótica y conclusión final obtenida

Categorías de gráfico construido	Conclusión			Total en el nivel
	Corr.	Parcial	Blanco o Inc.	
C1. Representa sólo sus datos			2	2
C2. Representa resultados individuales		3	12	15
C3. Gráficos separados para cada muestra	1	12	33	46
C4. Gráficos conjuntos	1	7	17	25
Total	2	22	64	88

Figura 4.16. Clasificación de estudiantes, según nivel de complejidad semiótica y conclusión final obtenida





Conclusiones erróneas

Así, algunos alumnos no perciben la impredecibilidad de los experimentos aleatorios, identificando ñbuena intuiciónò con capacidad de obtener exactamente los mismos resultados en el experimento real y simulado, es decir mostrando la ilusión de control descrita por Langer (1975), como se observa en los siguientes ejemplos:

ñLas conclusiones obtenidas después de realizar el experimento aleatorio de lanzar una moneda podemos comprobar el número de veces que la previsión coincide con la real para comprobar las diferencias entre las dos situacionesò (Alumno AA).

ñLas conclusiones a las que llego después de haber hecho el estudio acerca de las intuiciones de los estudiantes de la clase sobre el experimento de lanzar una moneda son: mis compañeros han tenido una buena intuición a la hora de realizar la secuencia simulada, llegando a obtener resultados idénticos entre las dos secuencias, como me ocurrió a miò (Alumno GC).

ñNo se parecen mucho ya que en las simuladas lo que se hacía es inventarse los datos y no sabes como pueden salir los datos en la realidad. Cada tirada tienes un 50% de posibilidades para acertarò (Alumno AG).

Otra concepción errónea sobre los fenómenos aleatorios la muestra el alumno EL quien no entiende el propósito de la estadística pues no comprende que, aunque el fenómeno aleatorio no sea predecible para el caso particular, la estadística permite detectar las tendencias y medir la variabilidad de los resultados en un conjunto de experimentos. De este modo, las similitudes en los resultados de las dos distribuciones las atribuye al azar y no a una buena intuición del conjunto de la clase.

ñEn mi opinión creo que los resultados han sido obtenidos por el azar. Comparando los datos me he dado cuenta que son muchos los compañeros que coinciden pero aún así sigo pensando que es mera casualidad, porque en la simulada hemos puesto lo que hemos querido puesto que nadie sabía lo que iba a salirò (Alumno EL).

Todas las conclusiones obtenidas a nivel C1 de complejidad son incorrectas así como la mayoría de las producidas a nivel C2; estos tipos de gráficos no contribuyen a finalizar la resolución del problema. La proporción de conclusiones incorrectas disminuye en los niveles superiores aunque aún es muy alta en todos los niveles.

Conclusiones parcialmente correctas

Veintidós estudiantes llegan a una conclusión parcial indicando 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. Vemos a continuación algunos ejemplos; en todos los casos se hace referencia a los promedios, a pesar que al interpretar las gráficas algunos estudiantes habían comparado las medidas de dispersión. Sin embargo estos estudiantes no asocian la diferencia de dispersión en las distribuciones real y simulada con una pobre intuición sobre la variabilidad aleatoria en los estudiantes.

ñPodemos concluir diciendo que las intuiciones de los estudiantes de la clase sobre el experimento aleatorio son muy acertadas y aproximadas a la realidad, aunque en las secuencias el número de caras es más exacto que el tamaño de las rachasò (Alumno JC).

ñEn general si hay una buena intuición ya que las medias son valores muy próximos, es decir han salido casi los mismos valores y a golpe de vista se puede comprobarò (Alumno TG).

ñLas intuiciones de la clase haciendo la media es bastante buena. Al calcular la moda, es decir el número que más se repite en el caso de la variable ñnúmero de carasò es el mismoò (Alumno CG).

Conclusiones correctas

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. A continuación reproducimos la respuesta de uno de ellos.

ñEn cuanto al número de caras las intuiciones del aula fueron aproximadas a la realidad, pero no del todo, ya que la desviación típica nos indica que los datos se distancianò. Anteriormente la alumna dice: ñLa media entre la simulada y la real se asemeja,.... la mediana y la moda dan los mismos datosò (Alumna CG)

Resaltamos el hecho de que en los dos casos, el gráfico producido por el estudiante es de nivel C4. Este tipo de gráfico ha permitido la solución completa al problema que no se produce en los niveles anteriores.

4.5.CONCLUSIONES DEL ESTUDIO EMPÍRICO

Como conclusiones del análisis de los gráficos producidos por los estudiantes destacamos, en primer lugar, el interés de llevar a cabo un estudio semiótico de los mismos puesto que ello nos permite mostrar que diferentes representaciones gráficas del mismo conjunto de datos no son equivalentes en cuanto a la configuración de objetos matemáticos que el alumno pone en juego en su construcción.

La jerarquía iniciada de niveles de complejidad semiótica parece pertinente, pues aunque el número de estudiantes participantes es moderado y tan sólo analizamos un proyecto de análisis de datos, los resultados parecen indicar que un nivel superior en la jerarquía incide en el nivel de lectura y comprensión del gráfico producido así como en la obtención de conclusiones correctas al menos parcialmente.

Una segunda conclusión del estudio que es independiente del interés de relacionar la complejidad del gráfico con su interpretación y conclusión es el hecho preocupante que un número tan pequeño de estudiantes llegue a una conclusión completa en el proyecto planteado. En dicho proyecto se pretende que los estudiantes recorran 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) ñ*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"* (pg. 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, 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 nos parece necesario que los futuros profesores comprendan y puedan llevar a cabo todos los pasos de dicho proceso en actividades similares a la analizada. Además, el trabajo con proyectos abiertos puede ser especialmente adecuado en el trabajo individual y en grupos recomendados en el marco del Espacio Europeo de Educación Superior, y por ello pensamos que debieran emplearse en la formación de profesores.

Nuestra investigación también indica que la construcción e interpretación de gráficos (incluso elementales) 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.

En nuestros alumnos se confirman algunos errores frecuentes en la construcción de gráficos citados por dichas autoras, ya que pocos estudiantes incluyen un rótulo correcto y significativo en el gráfico y algunos no centran los intervalos de frecuencias en los histogramas. Hemos encontrado errores no descritos por Bruno y Espinel como representar el producto de los valores de la variable por su frecuencia y cambiar la variable dependiente e independiente, representando las frecuencias en el eje de las X, error ya señalado en el trabajo de Ruiz (2006).

5. CONCLUSIONES

En este trabajo hemos planteado a los estudiantes de Magisterio una actividad abierta, que nos ha permitido iniciarnos en la investigación, realizando un primer estado de la cuestión, identificando un problema sencillo de investigación, recogiendo y analizando los datos y obteniendo unas primeras conclusiones. Como hemos indicado, este trabajo se incluye en la línea de investigación sobre Didáctica de la Estadística en la que se trabaja en el Departamento de Didáctica de la Matemática, concretándonos en la evaluación del conocimiento de los futuros profesores de educación primaria respecto a los gráficos estadísticos.

En lo que sigue expondremos las conclusiones obtenidas con respecto a los objetivos y las hipótesis y finalizaremos con limitaciones encontradas en nuestro estudio y futuras líneas de investigación.

5.1. CONCLUSIONES SOBRE LOS OBJETIVOS

A continuación vamos a exponer las conclusiones a las que hemos llegado con respecto a los objetivos que se expusieron en el apartado 2.3 de este trabajo:

Objetivo 1. Iniciar un estado de la cuestión de las investigaciones sobre comprensión de gráficas estadísticas, especialmente las que se relacionan con la formación de profesores.

Este objetivo se ha conseguido parcialmente, ya que la breve síntesis de las investigaciones que hemos localizado y resumido en el capítulo 3 lo que constituye es un primer paso para un futuro estado de la cuestión más completo sobre la comprensión de gráficos estadísticos. Además, como ya nos propusimos al inicio de esta investigación, deberemos analizar con más detalle las investigaciones relacionadas con la comprensión gráfica y la formación de profesores, en particular las que se han presentado sobre este tema en la reciente conferencia del Joint ICMI/IASE Study (http://www.ugr.es/~icmi/iase_study/) ya que en la primera revisión de la literatura no hemos encontrado demasiados trabajos que traten esta problemática.

Objetivo 2. Identificar un problema de investigación de interés dentro de la formación de profesores y en el campo de la estadística.

En nuestro trabajo nos hemos interesado por la construcción de los gráficos estadísticos por parte de futuros profesores de Educación primaria. Pensamos que la línea de investigación iniciada tiene interés, tanto desde el punto de vista de la educación estadística, como en el campo de formación de profesores. En Educación estadística, porque se aborda el tema de los niveles en la comprensión gráfica desde un nuevo punto de vista: centrándonos en la construcción, en lugar de en la lectura y utilizando ideas del Enfoque Ontosemiótico para llevar a cabo la definición de niveles de complejidad del gráfico.

Respecto a la formación de profesores, como hemos dicho, en la revisión inicial son muy pocas las investigaciones encontradas sobre comprensión o construcción de gráficos estadísticos por futuros profesores. Por ello, nuestro trabajo puede proporcionar información útil al formador de profesores sobre las dificultades de los futuros profesores con la construcción de gráficos y el trabajo con proyectos.

Objetivo 3. Explorar algún instrumento de obtención de datos de futuros profesores sobre su capacidad de construcción de gráficos.

Son muchos los posibles intrumentos para recoger datos sobre esta capacidad, por ejemplo, cuestionarios o entrevistas. Entre los posibles instrumentos de obtención de datos elegimos un proyecto abierto de análisis de datos realizado por futuros profesores de Educación primaria de la Universidad de Granada. Este mismo instrumento ha sido también usado en otras investigaciones realizadas en el grupo de investigación (como las de Tauber, 2001 y Alvarado, 2007). Esta elección está justificada ampliamente en el capítulo 2, en el apartado 2.6., donde hemos analizado como el trabajo con proyectos es deseable para la formación estadística de los futuros profesores y el trabajo con actividades de este tipo se incluye en la Educación Primaria en los Decretos de Enseñanzas Mínimas. Mediante el trabajo con proyectos, no sólo obtenemos los datos para la investigación sino contribuimos a preparar al futuro profesor para abordar la enseñanza de la estadística con esta metodología.

Objetivo 4. Recoger algunos gráficos elaborados por futuros profesores con dicho instrumento y clasificarlos para definir una jerarquía inicial de niveles de complejidad en la construcción de dichos gráficos.

Este era el objetivo principal de nuestra investigación y creemos haberlo llevado a cabo. En el capítulo 4 se han analizado los gráficos realizados por los futuros profesores en el proyecto de análisis de datos ñComprueba tus intuiciones sobre el azarò y se han definido unos niveles de complejidad semiótica, que nos han permitido clasificar los gráficos elaborados por la muestra de futuros profesores de nuestro estudio. En este sentido aportamos un resultado original, ya que no hemos encontrado en nuestra revisión bibliográfica ninguna categorización basada en la construcción de los gráficos por parte de los alumnos.

5.2. CONCLUSIONES SOBRE LAS HIPÓTESIS

En el capítulo 2 se definieron dos hipótesis de investigación y a continuación pasamos a comentar las conclusiones obtenidas respecto a dichas hipótesis:

Hipótesis 1. Los futuros profesores de nuestra muestra, al construir los gráficos estadísticos para llevar a cabo la tarea propuesta, cometen muchos de los errores ya detectados en investigaciones sobre dificultades y errores en la construcción de los gráficos estadísticos.

Después de realizar el análisis de los gráficos de los alumnos de nuestra muestra llegamos a la conclusión de que esta hipótesis se ha cumplido totalmente. La mayoría de los errores de los cuales nos alertan las investigaciones previas aparecen reflejados en los gráficos construidos por los futuros profesores participantes en nuestro estudio. Así hemos encontrado que los alumnos no centran los intervalos al construir histogramas; no unen con el eje X los polígonos de frecuencias, olvidan los rótulos o hacen rótulos confusos (Lee y Meletiou, 2003; Bruno y Espinel; 2005; Espinel, 2007;), confunden la variable dependiente e independiente en la distribución de frecuencias (Ruiz, 2006) y hacen un uso acrítido del software cuando lo utilizan para construir sus gráficos (Ben-Zvi y Friedlander, 1997).

Además en nuestro análisis hemos encontrado errores no descritos en las investigaciones previas como representar el producto de los valores de la variable por su frecuencia o representar sólo sus datos individuales.

Hipótesis 2. Los dos niveles superiores de complejidad semiótica definidos en nuestra investigación no son alcanzados por la mayoría de los estudiantes de la muestra de nuestro estudio.

Esta hipótesis, al realizar el estudio empírico, vemos que no se verifica, como ya se mostró en el capítulo 4 de este trabajo, puesto que el 80,6% del total de los alumnos que realizan gráficos lo hacen dentro de una de las dos categorías superiores (niveles C3 y C4). Es decir, la mayoría de los estudiantes de nuestra muestra llegan al concepto de distribución de una variable estadística, que según Wild y Pfannkuch (1999) es parte esencial del razonamiento estadístico. Esto nos muestra que la mayoría de los futuros profesores llegan a construir un gráfico de complejidad suficiente para poder resolver con éxito el problema que se les planteaba en el proyecto ñComprueba tus intuiciones sobre el azarò.

Los problemas se presentan porque, aunque el gráfico sea suficientemente complejo, bien es inadecuado (por ejemplo se representa en el mismo gráfico variables no relacionadas, como el número de caras y el número de rachas) o bien los alumnos se limitan a construir el gráfico pero no lo interpretan o lo interpretan incorrectamente. Hay también dificultades a la hora de relacionar la interpretación matemática del gráfico con la pregunta planteada sobre las intuiciones.

5.3. LIMITACIONES DEL ESTUDIO Y FUTURAS LÍNEAS DE INVESTIGACIÓN

El proyecto del cual hemos obtenido los datos para nuestro estudio formaba parte de una práctica más amplia cuyos objetivos principales eran adquirir competencias de análisis didáctico de experiencias de enseñanza y también competencias de búsqueda de recursos didácticos. En este trabajo sólo nos hemos centrado en analizar los gráficos estadísticos construidos como parte de la actividad desde distintos puntos de vista. Además, en este proyecto los alumnos trabajaron y analizaron tres pares de variables estadísticas, nosotros en este caso nos hemos centrado en el estudio de dos de ellas, el trabajo sería más completo si próximamente estudiamos el análisis que realizan los estudiantes de todas las variables puestas en juego en el proyecto.

Otra limitación de nuestro trabajo es el tamaño de la muestra que podría ser ampliado en un futuro. Además en esta investigación analizamos solo los resultados de los alumnos en una única tarea. En un futuro podrían incluirse distintas tareas, por ejemplo alguna en la que los alumnos debieran interpretar críticamente gráficos estadísticos dados por el profesor, en la línea de las investigaciones de Aoyama (2007) y Monteiro y Ainley (2007). En nuestro trabajo hemos estudiado las interpretaciones de los alumnos pero de gráficos construidos por ellos mismos, en este sentido una conjetura que hemos formado a partir de los resultados es que cuanto menor sea el nivel de complejidad del gráfico más fácil le resulta al alumno interpretarlo, que podríamos tratar de confirmar.

En un futuro se podría ampliar este estudio analizando el conocimiento didáctico de los profesores para enseñar estadística y en particular los gráficos estadísticos. Este conocimiento lo podríamos evaluar a partir de las unidades didácticas producidas por ellos para enseñar gráficos estadísticos, de su evaluación y clasificación de tareas relacionadas con los gráficos (en la línea del trabajo de González y Pinto, 2008) o de la evaluación que hacen de gráficos producidos por estudiantes (siguiendo a Espinel, 2007).

6. REFERENCIAS

Alvarado, H. (2007). *Significados institucionales y personales del teorema central del límite en la enseñanza de estadística en ingeniería.* Tesis Doctoral. Universidad de Granada.

Arteaga, P., Batanero, C. y Ruiz, B. (En prensa). Complejidad semiótica de gráficos estadísticos en la comparación de dos distribuciones por futuros profesores. En Hipótesis Alternativa, número monográfico sobre el Encuentro Latino Americano de Educación Estadística (ELEE, Monterrey, México 4-5 Julio, 2008).

Arteaga, P., Ortiz, J. J. y Batanero, C. (2008). Uso de gráficos comparación de dos distribuciones por futuros profesores. Poster presentado en *RELME 22,* México, Julio, 2008.

Arteaga, P., Batanero, C., Ortiz, J. J. y Ruiz, B. (2008). Complejidad semiótica de gráficos estadísticos en la comparación de dos distribuciones por futuros profesores. Trabajo presentado en el Grupo de Probabilidad y Estadística. *XIII Simposio de la Sociedad Española de Educación Matemática.* Badajoz.

Aoyama. K. (2007). Investigating a hierarchy of studentsô interpretations of graphs. IEJME

Aoyama, K. 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. G. Burrill y M. Camden (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/.

Batanero, C. (2001). *Didáctica de la Estadística.* Granada: Grupo de Investigación en Educación Estadística. Didáctica de las Matemáticas.

Batanero, C., Arteaga, P. y Díaz, C. (En prensa). El lenguaje de los gráficos estadísticos. Investigaciones sobre su comprensión. En *Libro Homenaje a Antonio Romero*. Departamento de Didáctica de la Lengua. Universidad de Granada.

Batanero, C. y Díaz, C. (2004). El papel de los proyectos en la enseñanza y aprendizaje de la estadística. En J. Patricio Royo (Ed.), *Aspectos didácticos de las matemáticas* (125-164). Zaragoza: ICE.

Ben-Zvi, D., y Friedlander, A. (1997). Statistical thinking in a technological environment. En J. Garfield y G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 54-64). Voorburgo, International Statistical Institute.

Bertin (1967). *Semiologie graphique*. Paris: Gauthier-Villars.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review, 65*(2), 167-190.

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 VII*, 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.

Connor, D., Davies, N. y Payne, B. (2002). Web-based project and key skill work. *Teaching Statistics*, *24*(2), 62-65.

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

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

delMas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study studentsô difficulty reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Proceedings of the 4th International Research Forum on Statistical Reasoning, Thinking and Literacy*. University of Auckland. Online: app.gen.umn.edu/artist/articles/SRTL4_ARTIST.pdf

Diaz, C. y Arteaga, P. (2008). Developing general and data analysis competences through project work. Trabajo presentado en el *III European Congress of Methodology*, Oviedo, 2008.

Díaz, C., Arteaga, P. y Batanero, C. (2007). Contribución del trabajo con proyectos estadísticos a la adquisición de competencias básicas. En M. Molina, P. Pérez-Tyteca y M. A. Fresno (Eds.). *Jornadas de Investigación en el Aula de Matemáticas. Competencias matemáticas*. Granada. Sociedad Thales y Departamento de Didáctica de las Matemáticas. CD- ROM.

Estepa, A. (1993). *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.

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

Espinel, C., Bruno, A. y Plasencia, I. (2008). Statistical graphs in the training of teachers. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.* Monterrey: ICMI e IASE. CD-ROM.

Font, J. D., Godino, J. D. y DôAmore, B. (2007). An ontosemiotic approach to representations in mathematics education. *For the Learning of Mathematics, 27* (2), 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), 124-158.

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

Gerber, R., Boulton-Lewis, G y Bruce, C. (1995). Childrenôs understanding of graphic representation of quantitative data. *Learning and Instruction 5,* 70-100.

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

Godino, J. D., Batanero, C. y Fonts, V. (2007). The onto-semiotic approach to research in mathematics education. *Zentralblatt für Didaktik der Mathematik, 39*(1-2), 127-135.

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 C. Batanero, G. Burrill, C. Reading y A. Rossman (2008). *Proceedings of the Joint ICMI /IASE StudyTeaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey, México: ICMI e IASE. CD ROM, en prensa.

González, M. T. y Pinto, J. (2008). Conceptions of four pre-service teachers on graphical representation. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.* Monterrey: ICMI e IASE. CD-ROM.

Graham, A. (1987). *Statistical investigations in the secondary school.* Cambridge: The

Open University Centre for Mathematics Education.

Henry, M. (1997). Notion de modéle et modélization en lõenseignement. En *Enseigner les probabilités au* lycée (pp. 77-84). Reims: Commission Inter-IREM-

Holmes, P. (1997). Assessing project work by external examiners. En I. Gal y J. B: Garfield (Eds.), *The assesment challenge in statistics education* (pp. 153-164). Voorburg: IOS Press.

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

Li, D. Y. y Shen, S. M. (1992). Studentsõweaknesses in statistical projects. *Teaching Statistics 14* (1), 2-8.

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

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

MEC (2006b). *Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria.*

Monteiro, C. y Ainley, J. (2006). *Student teachers interpreting media graphs*. En A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil:* International Statistical Institute and International Association for Statistical Education. Online: http://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), 188-207. On line: http://www.iejme/.

Murray, S., y Gal, I. (2002). Preparing for diversity in statistics literacy: Institutional and educational implications. En B. Phillips (Ed.). *ICOTS-6 papers for school teachers.* Cape Town: International Association for Statistics Education (CD ROM).

Nolan, D., y Speed, T.P. (1999). Teaching statistics theory through applications. *American Statistician, 5*3, 370-375.

Pereira-Mendoza, L. y Mellor, J. (1990). Studentõs concepts of bar graph: Some preliminary findings. *Proceedings of the Third International Conference on*

Teaching Statistics Ed D. Vere-Jones. Voorburg: International Statistical Institute.

Pimenta, R. (2003). *O uso da estatística no projecto de investigação em fisioterapia*. Trabajo de Investigación Tutelada. Universidad de Santiago de Compostela.

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

Schield, M. (2006). Statistical literay survey analysis: reading graphs and tables of rates percentages. En B. Phillips (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

Starkings, S. (1997). Assessing studentsô projects. En I. Gal y J. B. Garfield (Eds.), *The assesment challenge in statistics education* (pp. 139-152). Voorburg: IOS Press.

Tauber, L. (2001). *Significado y comprensión de la distribución normal a partir de actividades de análisis de datos.* Tesis Doctoral. Universidad de Sevilla.

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