

DESARROLLO, VALIDACIÓN Y EVALUACIÓN DE UN CUESTIONARIO DE OPCIONES MÚLTIPLES PARA IDENTIFICAR Y CARACTERIZAR LAS VISIONES SOBRE LA NATURALEZA DE LA CIENCIA DE PROFESORES EN FORMACIÓN

MARÍN, NICOLÁS¹ y BENARROCH, ALICIA²

¹ Departamento de Didáctica de las Matemáticas y de las Ciencias Experimentales. Universidad de Almería.

² Departamento de Didáctica de las Ciencias Experimentales. Universidad de Granada.

nicolas.marin@gmail.com

aliciabb@ugr.es

Resumen. En este trabajo se muestra el diseño de un cuestionario de opciones múltiples (COMVDC) para evaluar creencias sobre el conocimiento de ciencias en el contexto de la formación de futuros profesores de secundaria. El proceso seguido ha implicado varios años de trabajo y ha conducido a un cuestionario con ventajas e inconvenientes respecto a otros instrumentos con objetivos similares. La característica más destacada de este proceso es que ha posibilitado la discriminación de ítems de baja calidad frente a otros de media y alta, convirtiéndose por tanto en un proceso de autocorrección. Se discute la validez y la fiabilidad del cuestionario en el estado que se presenta y se extraen conclusiones referidas a sus resultados y sus posibilidades para ser introducido como herramienta tanto de evaluación como de enseñanza en la formación de futuros profesores de ciencias.

Palabras clave. Creencias sobre ciencias, formación de profesores, visión de ciencia, cuestionario opciones múltiples.

Development, validation and evaluation of a multiple-choice questionnaire to identify and to characterize preservice teachers' science views

Summary. The design of a multiple-choice test (COMVDC) to evaluate scientific beliefs of future secondary teachers is shown. The followed way has involved several years of fieldwork and has ended in a questionnaire with advantages and disadvantages with respect to other instruments with similar objectives. The most outstanding characteristic of this process is that it has made possible the discrimination of questions of low quality as opposed to others of average and high, becoming therefore a self-correction process. We discuss the validity and the reliability of the current questionnaire and we extract conclusions about their possibilities to be introduced like a tool of evaluation and teaching in the formation of future teachers of science.

Keywords. Scientific beliefs, formation future teachers, science vision, questionnaire multiple options.

1. INTRODUCCIÓN

Una de las líneas de investigación que más fructífera se ha mostrado en los últimos años ha sido la de las concepciones, visiones, actitudes, etc., sobre naturaleza y construcción del conocimiento de ciencias. Según Lederman, Abd-El-Khalick, Bell y Schwartz (2002), durante los pasados 90 años, casi todos los científicos, profesores de ciencias y organizaciones implicadas en la enseñanza de las ciencias han coincidido en la importancia de ayudar a los estudian-

tes a desarrollar concepciones adecuadas sobre la ciencia. Sin entrar a discutir el grado de exactitud de esta afirmación, que creemos algo exagerada, al menos en el ámbito que nos rodea, lo que sí parece claro es que las reformas acometidas recientemente en la educación científica han dado una importancia cada vez mayor al objetivo de formar adecuadamente a los estudiantes sobre la naturaleza misma de la ciencia (American Association for the Ad-

vancement of Science, 1990, 1993; National Research Council, 1996; OEI, 2001; CECJA, 1992, 1994, 1995). Lógicamente esta tendencia curricular es consecuencia y a su vez justifica la proliferación de trabajos en esta línea de investigación.

Paralelamente, se han hecho y se siguen haciendo (Aikenhead, 1988; Lederman, Wade y Bell, 1998) grandes esfuerzos para medir el grado de comprensión de estudiantes y profesores sobre la naturaleza de la ciencia. La mayoría de estos esfuerzos han usado instrumentos de lápiz y papel; generalmente tests estandarizados, que han sido repetidamente cuestionados por su escasa validez y fiabilidad. La falta de precisión en el objeto que se quiere medir (naturaleza de la ciencia), la naturaleza de los constructos implicados (concepción, visión, creencia, actitud, etc.), y la multidimensionalidad de estos constructos son algunos de los aspectos problemáticos que subyacen en estas mediciones. Concretamente, dado el carácter dialéctico y en muchos casos ausencia de especificación de los supuestos filosóficos que subyacen en la misma naturaleza de la ciencia, se debilitan e invalidan los datos e interpretaciones obtenidas mediante los instrumentos contruidos. Esto da lugar a uno de los errores y defectos más graves y más difíciles de valorar y corregir, como es la inadecuación de criterio, es decir, la falta de ajuste o correspondencia entre el método o instrumento elegido para la medida y el objeto que se quiere medir (Manassero y Vázquez, 2002). Otros problemas señalados son la tendencia de los alumnos a responder para satisfacer las expectativas del profesor y la dificultad del lenguaje técnico involucrado que deriva en problemas semánticos de significado y comprensión de la terminología empleada. Aikenhead (1988) estudió la ambigüedad del lenguaje contrastando las respuestas escritas obtenidas de distintos instrumentos de evaluación (Likert, párrafos escritos, elección múltiple derivada empíricamente y entrevistas semiestructuradas) con las respuestas en otras entrevistas complementarias. Concluyó que las entrevistas semiestructuradas reducen la ambigüedad casi a cero, pero la gran cantidad de tiempo necesario para recoger y analizar sus datos, le lleva a concluir que los cuestionarios de opción múltiple derivados empíricamente son el método más operativo para reducir la ambigüedad.

En el contexto de esta línea de investigación se desarrolla este trabajo, cuyo propósito es describir el desarrollo de un nuevo instrumento para evaluar creencias científicas en el contexto de la formación de futuros profesores de secundaria. Esto es, nuestro objetivo principal es que el cuestionario sea representativo de los contenidos que sobre la naturaleza de la ciencia se tratan de enseñar, a fin de que sus mediciones permitan la evaluación del grado de aprendizaje experimentado. Específicamente, en este trabajo pretendemos:

- a) Mostrar los contenidos y temáticas sobre la construcción y naturaleza de la ciencia que podrían ser de interés para futuros profesores de ciencias de secundaria, y que, en consecuencia, se pretenden evaluar con el cuestionario.
- b) Describir el proceso y ejemplificar la elaboración inicial de los ítems de opciones múltiples, atendiendo a la sistemática de contenidos anteriormente establecida.

c) Mostrar el proceso utilizado en la validación del cuestionario inicial, tras su administración a una muestra de 45 futuros profesores de secundaria en un curso del CAP desarrollado en 2003.

d) Discutir la utilidad y las limitaciones del cuestionario final, tras su administración a una segunda muestra de 50 futuros profesores de secundaria en un curso del CAP desarrollado en 2005.

e) Extraer conclusiones referidas a los resultados del cuestionario y sus posibilidades para ser introducido como herramienta tanto de evaluación como de enseñanza en la formación de futuros profesores de ciencias.

2. REVISIÓN DE LOS PRINCIPALES INSTRUMENTOS USADOS PARA EVALUAR VISIONES DE CIENCIAS

En la bibliografía sobre la temática que nos ocupa, se han usado términos como «visiones de ciencias» (McComas, Clough y Almazroa, 1998; Fernández, Gil, Carrascosa, Cachapuz y Praia, 2002), «ideas sobre ciencias» (Osborne, Collins, Ratcliffe, Millar y Duschl, 2003) o «naturaleza de la ciencia». Aun sabiendo que este último es el más consensuado en la literatura y cada vez más se extiende su uso en su sentido más amplio (Vázquez, Acevedo y Manassero, 2004), utilizaremos con más frecuencia el término *visión de ciencias* –y su acrónimo *VdC*– puesto que creemos que recoge mejor los aspectos epistemológicos de la actividad constructiva de este conocimiento.

A continuación, se revisan, por orden cronológico, algunos de los instrumentos más conocidos usados en la medición de concepciones, visiones, creencias o actitudes científicas:

a) El inventario *Views on Science-Technology-Society (VOSTS)*, preparado por Aikenhead, Fleming y Ryan (1987) y modificado posteriormente por Aikenhead y Ryan (1992). Estos autores explicaron los pasos que siguieron para su desarrollo y las discusiones sobre su validez y fiabilidad. El VOSTS está formado por 114 ítems de opciones múltiples cuya principal característica es que están elaborados empíricamente, es decir, las opciones recogen las opiniones expresadas en entrevistas piloto anteriores o las respuestas a cuestiones abiertas dadas por personas semejantes a las que está destinado el cuestionario. De este modo, estos nuevos instrumentos constituyen una vía de evaluación intermedia, al intentar aunar las ventajas de los cuestionarios (por ejemplo, su agilidad en la aplicación) con las de las entrevistas (por ejemplo, su fiabilidad y precisión) (Acevedo y otros, 2001).

– El cuestionario VOSTS abarca temas correspondientes a las siguientes dimensiones conceptuales: definiciones de ciencia y tecnología, interacciones entre ciencia, tecnología y sociedad, sociología externa de la ciencia, sociología interna de la ciencia (características de científicos, construcción social de la tecnología y del conocimiento científico, ciencia y género) y naturaleza del conocimiento científico.

El cuestionario, según sus autores, presenta ventajas métricas destacables (fidelidad de las medidas y capacidad de discriminación de respuestas), pero ven el inconveniente de que no se usa toda la información de las opciones dado que el entrevistado elige sólo una y, por tanto, se desconoce qué opina sobre las demás.

b) Las limitaciones del Modelo de Respuesta Única (MRU) usado por sus autores en el VOSTS hacen que sólo pueda ser utilizado para hacer comparaciones centradas en cada cuestión particular y no para hacer las comparaciones test-retest o las pruebas de verificación de hipótesis, básicas de la estadística inferencial. Rubba y Harkness (1993) desarrollaron un conjunto de 16 ítems de opción múltiple denominado Teacher's Belief About Science-Technology-Society (TBA-STs) para investigar las creencias de los profesores sobre los temas CTS. En estos ítems, se usó un nuevo procedimiento de puntuación en el que las respuestas fueron clasificadas como adecuadas, plausibles o ingenuas, asignándoles valores de 3, 2 o 1 respectivamente a las categorías. Esta asignación fue realizada por un panel de jueces. Con este sistema, los autores pudieron evaluar los cambios globales que experimentaban las creencias científicas de los profesores tras la realización de cursos específicos sobre Ciencia-Tecnología-Sociedad (Schoneweg, Rubba y Harkness, 1995).

c) Paralelamente, Acevedo (1992; 1994) en nuestro país, diseñó un nuevo cuestionario tipo Likert denominado Cuestionario de Opiniones sobre Ciencia y Sociedad (COCS), que consta de 20 enunciados, expresados unos en términos positivos y otros de forma negativa, que se relacionan con seis grandes dimensiones derivadas de la sociología y la epistemología de la ciencia:

- Control sociopolítico de la investigación científica y tecnológica.
- Neutralidad ideológica de la ciencia y la tecnología.
- La objetividad como cualidad esencial de los científicos.
- Estereotipos sexistas en ciencia y tecnología.
- Creencias epistemológicas sobre la naturaleza del conocimiento científico.
- La ciencia como medio principal para la resolución de problemas sociales.

Este cuestionario fue administrado a estudiantes (Acevedo, 1992) y a futuros profesores (Acevedo, 1994). Los entrevistados debían expresar su grado de acuerdo con la afirmación mediante una cifra comprendida entre 1 y 6.

d) Desde una línea de trabajo diferente, concretamente la de la medición de actitudes relacionadas con la ciencia, Vázquez y Manassero (1997) elaboraron un Protocolo de Actitudes relacionadas con la Ciencia (PAC) formado

por 50 frases sencillas y cortas que eran valoradas sobre una escala de cinco puntos tipo Likert. Lo interesante del cuestionario es que estaba construido en base a una taxonomía de 4 dimensiones (dentro de las cuales se distinguían subdimensiones), a saber:

- Actitudes relacionadas con la enseñanza-aprendizaje de la Ciencia y Tecnología.
- Actitudes relacionadas con las interacciones entre la sociedad y la Ciencia y Tecnología.
- Actitudes relacionadas con la naturaleza del conocimiento científico y técnico.
- Actitudes relacionadas con los métodos y procedimientos científicos.

e) Sin embargo, como algún tiempo después estos mismos autores reconocen, «la simple suma de las valoraciones asignadas (como hacen las escalas Likert) no es una buena medida de las actitudes, pues las distintas alternativas tienen diferentes valores actitudinales» (Manassero y Vázquez, 2002), lo que les lleva a abandonar los cuestionarios con escala Likert y adoptar los de opciones múltiples; concretamente, el cuestionario VOSTS. De hecho en algunas revisiones, como la que realizan Lederman, Wade y Bell (1998), el VOSTS sale muy bien parado. Finalmente los autores del PAC, basándose en el VOSTS y en el TBA-STs, adaptando y refundiendo algunas de las cuestiones contenidas en estos cuestionarios, terminaron construyendo una versión en español denominada COCTS (Cuestionario de Opiniones sobre Ciencia, Tecnología y Sociedad) con 100 cuestiones (Vázquez y Manassero, 1998; 1999) que contienen un total de 637 frases donde se reflejan distintas creencias y actitudes sobre la naturaleza de la ciencia desde una perspectiva CTS más amplia. Para superar el inconveniente de la respuesta única usada en el VOSTS, en este nuevo formato se pide al entrevistado que valore el grado de acuerdo con todas y cada una de las opciones presentes en cada cuestión sobre una escala de nueve puntos. Además, mediante un proceso de validación, se categorizan las distintas alternativas como adecuadas, plausibles e ingenuas, al modo que se hiciera con el TBA-STs. No vamos a extendernos en el sofisticado proceso de métrica usado por sus autores para la evaluación de las actitudes con este instrumento, pues remitimos al lector a sus fuentes directas (Vázquez, Acevedo y Manassero, 2000; Vázquez, Manassero y Acevedo, 2005). El cuestionario completo se puede ver en:

<http://store.digitalriver.com/store/ets/DisplayProductDetailsPage/productID.39407800>

f) Un cuestionario de características distintas ha sido desarrollado por Lederman, Abd-El-Khalick, Bell y Schwartz (2002). Estos autores consideran que la falta de consenso que continúa existiendo entre filósofos, historiadores, sociólogos y educadores de la ciencia acerca de muchos aspectos de la naturaleza de la ciencia es irrelevante para la instrucción científica. Limitan su estudio a los siguientes contenidos:

- El conocimiento científico es tentativo, empírico, cargado de teoría, en parte producto de la inferencia humana, imaginación y creatividad, e influido por la sociedad y la cultura.
- Distinción entre observación e inferencia.
- Inexistencia de un método estándar para hacer ciencia.
- Relaciones entre teorías científicas y las leyes.

Por otro lado, tratando de evitar los problemas inherentes a los instrumentos de opciones múltiples, sobre todo los semánticos de significado y comprensión de la terminología empleada, estos autores proponen un cuestionario de respuesta abierta complementado con entrevistas. El proceso de validación, que fue realizado por contraste de las respuestas al cuestionario con las obtenidas en entrevistas, ha llevado a los autores a la revisión del instrumento que finalmente presentan en tres formas distintas (formas A-VNOS, B-NOS, C-NOS) y para lo cual ha sido administrado a más de 2.000 estudiantes y profesores en formación y en activo de cuatro continentes diferentes.

La conclusión más evidente que se puede extraer de esta breve revisión, por supuesto inacabada, es que el esfuerzo por diseñar un instrumento para medir creencias sobre la naturaleza de la ciencia ha sido encomiable. Sin duda, este esfuerzo ha ayudado notablemente a entender e interpretar el conocimiento de ciencias entre los expertos implicados en el área (McComas, Clough y Almazroa, 1998; Izquierdo, 2000; Fernández, Gil, Carrascosa, Cachapuz y Praia, 2002; Osborne, Collins, Ratcliffe, Millar y Duschl, 2003). Se puede decir que la naturaleza de la ciencia en sí misma como área de conocimientos multidisciplinar, dialéctica, compleja y cambiante muestra dos corrientes de opinión contrapuestas (Vázquez, Acevedo y Manassero, 2004): la de quienes opinan que no es posible alcanzar consensos mínimos sobre ella y la de quienes opinan que, a pesar de la complejidad señalada, es posible encontrar algunos islotes de consenso en un mar de disensos. Nuestra posición se alinea en la última de las opciones, de modo que estamos convencidos que, tras un proceso de análisis comparativo, es posible alcanzar consensos mínimos sobre lo que es la ciencia, y que además son esos contenidos mínimos los que nos han de servir para el diseño del currículo de los futuros profesores.

Por otro lado, entre las diferentes técnicas e instrumentos utilizados, preguntas abiertas, entrevistas semiestructuradas, cuestionarios de opciones múltiples, etc., son estos últimos los que se van imponiendo por sus mejores resultados, siempre que las opciones se formulen a partir de respuestas a preguntas abiertas (cuestionarios empíricamente contruados). Dentro de ellos, se distinguen dos modelos de respuesta: el modelo de respuesta única, donde se elige sólo una de las opciones, y el modelo de respuesta múltiple en el que cada opción se evalúa con una escala. Los resultados muestran que esta última opción es más válida (Vázquez, Manassero y Acevedo, 2005), principalmente porque el modelo de respuesta única no usa toda la información disponible en las alternativas no seleccionadas.

3. ¿POR QUÉ UN NUEVO CUESTIONARIO SOBRE VdC?

Nuestro interés se ha centrado en evaluar el grado de aprendizaje de los futuros profesores de ciencias de secundaria sobre varios aspectos relacionados con la construcción y naturaleza de la ciencia. Por tanto, buscamos responder a un problema de aula: evaluar el aprendizaje mediante un instrumento adaptado a las peculiaridades contextuales de estos estudios de formación. Concretamente, creemos que un buen cuestionario debe:

- Ser breve para poder ser aplicado en el contexto de clase sin limar mucho tiempo de docencia.
- Admitir una buena integración, como una herramienta didáctica más, en el proceso de enseñanza.
- Apreciar con claridad los incrementos de aprendizaje en los contenidos de enseñanza.
- Caracterizar grupos de alumnos atendiendo a su visión de ciencias.
- Permitir tratamientos estadísticos de baja complejidad con el fin de evaluar tanto la bondad de los ítems como las ideas, creencias y posiciones epistemológicas de los sujetos de la muestra.
- Evaluar de forma homogénea los distintos contenidos ligados a la VdC de modo que, por ejemplo, no desarrollara en exceso las cuestiones epistemológicas en detrimento de las cuestiones CTS o viceversa. A la vez, por razones obvias, el cuestionario debía contemplar primordialmente los contenidos de VdC que en la actualidad gozan de un mayor consenso.

A pesar de las bondades manifiestas de los cuestionarios VOSTS y COCTS, creemos que requieren ciertas adecuaciones para ser integrados en un contexto de formación de futuros profesores de ciencias de secundaria. El cuestionario pretendido debe considerar las características anteriormente señaladas y, por tanto, tiene que servir para medir las visiones de ciencias en el contexto de la formación del profesorado de secundaria, de modo que su aplicación pretest-postest pueda revelar el grado del aprendizaje de los futuros profesores sobre los contenidos relacionados con la naturaleza de la ciencia enseñada.

El instrumento finalmente construido ha sido denominado *Cuestionario de Opciones Múltiples sobre la Visión de Ciencias -COMVDC-*. En el anexo I puede verse en la versión que se administró en 2005. Todos sus ítems tienen la misma estructura: una base donde brevemente se plantea una cuestión o un problema y tres opciones como respuesta, dos de las cuales son respuestas alternativas (incorrectas o menos adecuadas) y una es la respuesta correcta (o la mejor respuesta). Así construido, el COMVDC muestra ventajas y desventajas frente a los referidos VOSTS y COCTS.

Veamos las principales ventajas:

– La ventaja más evidente es su brevedad. La administración es bastante más liviana y sólo requiere de una hora para ser respondido.

– Admite un registro de datos bastante eficaz y ágil a través de plantillas que, a su vez, permiten una lectura automatizada (plantilla de lector óptico).

– Por cálculos sencillos a partir de la matriz de datos, se llega a una serie de índices que permiten la evaluación de cada uno de los ítems y de la prueba en general. Veremos más adelante estos índices. Aunque esta evaluación también es posible en cuestionarios de respuestas múltiples, la complejidad es mayor.

– Permite caracterizar sujetos de tendencia constructivista, racionalista o empirista, mediante patrones más definidos que los cuestionarios de respuesta múltiple. Estos pueden dar incluso respuestas paradójicas para un mismo ítem lo que dificulta la caracterización del entrevistado y ofrece, en el mejor de los casos, patrones de respuestas más complejos y menos delimitados.

– Precisamente por ofrecer patrones de respuestas más definidos, el seguimiento de sujetos a través de pretest y postest se hacen de un modo más preciso, lo que hace que COMVDC sea muy adecuado para experiencias docentes de aula donde se quiera evaluar progresos de aprendizaje.

Existen también algunas desventajas del COMVDC que se deben precisar para paliarlas en la medida de lo posible. Veámoslas:

– La información obtenida mediante COMVDC es menor que la recabada con los cuestionarios de respuesta múltiple. En nuestro caso, se ha tratado de compensar esta debilidad tomando las siguientes medidas:

- Estructurando altamente la prueba, de modo que cada aspecto relacionado con la VdC sea evaluado con varios ítems diseñados desde distintas perspectivas. Así, cada ítem se convierte en una unidad mínima de elección o decisión, procurando que tanto su base como las opciones se polaricen sólo a ese aspecto o perspectiva específica que se pretende evaluar.

- Construyendo las opciones de cada ítem a partir de las respuestas bien documentadas de los cuestionarios de opciones múltiples de la bibliografía consultada.

– Puesto que en COMVDC se «obliga» a elegir sólo una opción, se puede suponer que la validez de la información obtenida resulta menor que en los cuestionarios de respuesta múltiple.

Ahora bien, la pérdida de información se revierte en simplicidad y posibilidad de evaluar cuantitativamente la calidad de cada ítem, lo cual permite depurar la validez, como explicaremos más adelante. Por otro lado, trabajando con una sistemática bien estructurada siempre es posible introducir nuevos ítems, partir de

un exceso de ellos y posteriormente depurar el cuestionario descartando los de menos calidad.

– Se puede achacar a COMVDC, como a todos los cuestionarios de respuesta única, que los sesgos de interpretación por parte del entrevistado sean mayores que con los cuestionarios de respuesta múltiple. En nuestro caso, se recoge una información adicional solicitando al entrevistado una evaluación del ítem sobre la bondad en la construcción del mismo (parámetros cuantitativos tipo V que veremos a continuación). Esta información va a permitir hacer una buena discriminación de ítems y, por tanto, reducir los sesgos de interpretación.

4. DISEÑO, DESARROLLO Y ADMINISTRACIÓN DEL CUESTIONARIO

A continuación, se va a mostrar el proceso de construcción del cuestionario según la secuencia temporal en que fue desarrollado:

– *Construcción de una sistemática de contextos sobre visiones consensuadas de ciencia.*

El punto de partida de esta investigación fue la elaboración de los aspectos acerca de la naturaleza de la ciencia que conforman un mínimo consensuado por el colectivo de investigadores del área. Esto se hizo a partir de trabajos realizados sobre visiones deformadas de ciencias (McComas, Clough y Almazroa, 1998; Fernández, Gil, Carrasco, Cachapuz y Praia, 2002; Osborne, Collins, Ratcliffe, Millar y Duschl, 2003; Marín, 2003; Vázquez, Acevedo y Manassero, 2004) y, aunque el proceso completo de comparación y síntesis será objeto de otro trabajo posterior, en la tabla 1 se recogen los resultados alcanzados.

Por ello, la tabla 1 se ha denominado «sistemática de contextos para diseñar el cuestionario sobre VdC», pues muestra la relación que se ha mantenido durante todo el proceso entre los contenidos o aspectos sobre la naturaleza de la ciencia que se pretenden evaluar y los ítems del cuestionario. Los contenidos metacientíficos se han agrupado en cuatro grandes categorías. Las tres primeras, referidas respectivamente: a) al contexto social y tecnológico en el que surge y se aplica la ciencia; b) al proceso investigador y c) a la regulación social de este proceso, están directamente relacionadas con la ciencia como proceso, mientras que la cuarta categoría se refiere a la naturaleza de la *ciencia como producto*. Dentro de cada categoría se distinguen subcategorías, contextos o aspectos concretos.

De este modo, los distintos contextos y temáticas relacionadas con la VdC se listan en la columna 2 de la tabla. En la columna 3 se expresa la visión más adecuada según los consensos alcanzados en la bibliografía consultada (Marín, 2003; Marín, 2005) y en la columna 4, la visión menos adecuada.

La base de datos de ítems de tres opciones se fue desarrollando en varios años; por un lado, partiendo de algunos

de los trabajos mencionados anteriormente y, por otro, realizando versiones simplificadas de diversos ítems del COCTS. Los ítems de la base de datos obedecen al plan trazado por la sistemática de contextos sobre la VdC establecida en la tabla 1. En la columna 5 se muestran los índices de la base de datos de los ítems seleccionados para formar parte del cuestionario y, entre paréntesis, se muestra la estructura del ítem. Algunos ítems, los descartados entre 2003 y 2005, aparecen en gris. Más adelante se explica con detalle el proceso de eliminación.

En la última columna se muestra el número de ítems para cada contexto de la VdC. Como se puede apreciar, la batería de ítems resultó formada, finalmente, por 54 cuestiones de opciones múltiples, distribuidas en 9, 9, 9, y 27 ítems respectivamente para cada una de los contextos sobre VdC. Es necesario señalar que la falta de proporcionalidad en la distribución se debe a que la cuarta categoría se refiere a la naturaleza de las ciencias como producto, donde, además de existir un consenso más definido, se desdobra en un mayor número de apartados, siendo cada uno de ellos de gran relevancia en la Didáctica de las Ciencias.

– *Fase de construcción de la base de datos de ítems sobre VdC, antes del año 2003.*

La elaboración de ítems de opciones múltiples es un proceso bien documentado (Lafourcade, 1972; Rodríguez Diéguez, 1980; Abad et al., 2006) que, además de la experiencia acumulada en más de dos décadas, se regula mediante los siguientes criterios y medidas:

– Uso de una misma estructura para todos los ítems, a saber: una base que consiste en una cuestión, un problema, una pregunta teórica, etc., y tres opciones como posibles respuestas, una de las cuales es correcta y las demás, llamadas distractores, son incorrectas o menos adecuadas.

– Las opciones que juegan el papel de distractores deben, al menos inicialmente, parecer tan buenas como la respuesta correcta. Con esto se pretende que el examinado utilice con precisión su conocimiento sobre el tema para hacer su elección y que ésta no venga «sugerida» por «estrategias» que no tengan nada que ver con los conocimientos que se desean evaluar. Por esta razón, se trata de evitar que la respuesta correcta sea más larga o más corta que los distractores, que todas las opciones armonicen gramaticalmente con la base y que tengan un grado de concreción semejante.

– Se debe tratar de formular cada ítem con la mayor precisión posible, de modo que una mayoría de expertos coincida en la respuesta elegida como más adecuada. Para ello, siempre que se ha podido, se han usado frases consensuadas en la bibliografía sobre el tema.

– Hay que favorecer que el evaluado entienda correctamente lo que se le pregunta, lo cual debe formularse de forma breve y, si es posible, expresado en afirmativo. Si existen palabras o frases que se repiten en todas las opciones, se intentan pasar a la base.

– La base debe evaluar un aspecto concreto tanto en comprensión como en extensión y las opciones giran sobre la base de tal modo que no evalúen otras cuestiones diferentes a la base. Con ello se pretende que para descartar una opción no se necesite otro conocimiento diferente al relacionado con la base.

– La elección debe ser un esfuerzo de discriminación cognitiva que no debe verse diluido por oraciones largas, de difícil lectura o incoherencias gramaticales. El mejor ítem está compuesto de oraciones cortas de fácil lectura.

Pero, además de los criterios anteriores, se aplicaron ciertas estrategias para intentar paliar algunas de las desventajas que se les suele achacar a los ítems de opciones múltiples de respuesta única. Una de las críticas que se suele hacer a estos ítems, es que sólo permiten tener en cuenta la información de la opción elegida por el sujeto. Este problema se ha compensado abordando cada contexto o aspecto ligado a la VdC con una diversidad de ítems.

Pero también es posible paliar parcialmente el problema prefijando el carácter de cada una de las opciones; por ejemplo, si la primera opción es una respuesta constructivista, la segunda racionalista y la tercera es empirista, tras la administración del cuestionario, se puede calcular las frecuencias de elección de cada opción y se puede obtener una información adicional de gran interés para el objetivo que se pretende, a saber, el control del aprendizaje del alumno sobre la naturaleza y visión de la ciencia.

De cada ítem hay dos versiones: la que usa el investigador en la base de datos, donde siempre la primera opción del ítem es la correcta, y la versión que se presenta al alumno, que es la que se presenta en el anexo 1, donde la respuesta correcta se «camufla» en cualquier lugar según una clave aleatoria, la misma que se utiliza en sentido inverso para asociar los índices.

Respecto a la versión del investigador, y a fin de ligar coherentemente el contenido teórico mostrado en la tabla 1 con el contenido específico de cada ítem, en su diseño y construcción se han considerado dos estructuras básicas:

– *Estructura constructivista (C).* a) la primera opción del ítem es una respuesta típicamente constructivista al planteamiento realizado en la base del ítem, b) la segunda opción debe ser más racionalista que la opción primera, y c) la tercera opción es más empirista que la opción primera. La primera opción, por ser la correcta, es coherente con las afirmaciones de la columna 3 de la tabla 1, y la segunda y tercera son coherentes con las realizadas en la columna 4. Los ítems así construidos se tipifican como C3. Un ejemplo puede verse en el ítem 26 del anexo I para el que la alternativa a) es la respuesta más empirista; la b) es la constructivista y la c) es la más racionalista. No siempre ha sido posible aplicar esta estructura a todos los ítems, bien porque la temática abordada por la base no lo permitía o resultaba difícil en exceso mantener la fórmula para cada opción, tal es el caso, por ejemplo, de los ítems 30, 31 y 32, que tratan el problema de la relación entre

la teoría y los datos empíricos. En este caso, al menos uno de los distractores, o los dos, ofrecen visiones empiristas. Se identifican como Ce.

- *Estructura organicista (O)*. Son ítems que abordan problemas ligados a las visiones más o menos complejas que se suele mantener sobre la construcción del conocimiento de ciencias. Se contemplan como visiones reduccionistas o simplistas que actualmente están claramente consensuadas (columna 4 de la tabla 1), a saber: n) visión descontextualizada, aproblemática y ahistórica de la ciencia; s) visión puramente racional del comportamiento de los científicos; r) visión rígida y algorítmica de la metodología científica; i) visión individualista de las aportaciones al cuerpo de conocimientos y a) visión acumulativa y lineal del crecimiento de la ciencia (para más detalle ver, por ejemplo, Fernández, Gil, Carrascosa, Cachapuz y Praia, 2002). Cuando el ítem aborda cuestiones generales sobre la relación entre teoría y realidad desde perspectivas organicistas y reduccionistas, entonces se tipifica como O3 (por ejemplo, en el ítem n.º 14 del anexo I. a) y b) son respuestas reduccionistas mientras que c) es una respuesta constructivista).

En todos los ítems de tipo O se respeta la siguiente estructura: a) la primera opción del ítem, la correcta, ofrece una respuesta típicamente organicista coherente con las afirmaciones de la columna 3 de la tabla 1 y b) la segunda y tercera opción del ítem deben ser respuestas claramente reduccionistas y, por tanto, coherentes con las afirmaciones de la columna 4 de la tabla 1.

- *Primera administración del cuestionario: año 2003.*

Un cuestionario inicial (constituido por 50 ítems de los 54 de la batería) fue administrado a una muestra de 45 futuros profesores de secundaria en enero del 2003. Junto al cuestionario, se proporcionó a los alumnos una plantilla para lector óptico para cumplimentar sus respuestas.

Cada ítem se responde de dos maneras: la primera, señalando la respuesta que se cree correcta; la segunda, valorando, entre cuatro opciones, la bondad del ítem desde la perspectiva del encuestado. Esta valoración permitió, como veremos en el apartado siguiente, calcular unos índices relativos a la calidad del ítem que fueron muy útiles para la selección final de los mismos. La valoración de la bondad del ítem se hace del siguiente modo:

- Si se ha respondido con seguridad, se marca la (a). No se marca en caso de que no haya existido esa seguridad.
- Si el ítem no parece bien construido, entonces se marca:
- b) si parece que entre las opciones no está la respuesta correcta.
- c) si parece que hay más de una respuesta correcta.
- d) si es que el ítem parece incoherente por motivos diferentes a b) o c).

En cualquier caso, pero en especial si se ha elegido b), c) o d), se puede realizar un comentario al ítem. Concretamente:

- Si la respuesta es b) se puede formular la opción que sería correcta.
- Si la respuesta es c) se puede indicar qué opciones parecen semejantes o proponer cómo cambiarlas.
- Si la respuesta es d) se pueden dar razones de por qué parece incoherente.

Una vez administrado el cuestionario, se pasa a una fase donde éste se usa en clase como un recurso didáctico para crear aprendizaje sobre cuestiones referentes a la VdC. Concretamente, la primera vez que se pasa en el 2003, se realiza una actividad en grupo donde tratan de ponerse de acuerdo respecto a las opiniones individuales sobre la bondad de cada ítem. Esto permite dar nuevos informes más objetivos sobre la calidad de los ítems.

Además, después de la experiencia docente, al finalizar el curso, se pasó una vez más el cuestionario, en este caso para que los sujetos expresaran su reacción afectiva al escuchar la respuesta elegida por los expertos. En el apartado siguiente se muestra cómo se calculan cinco nuevos índices (Sa, Sb, Sc, Sd y Se).

De este modo, para cada ítem se calculan 16 índices, 6 de ellos a partir de la respuesta del alumno al contenido del mismo, 5 relativos a la valoración que hacen de su bondad y 5 sobre la reacción afectiva al escuchar la respuesta correcta.

Mediante estos índices, y siguiendo una dinámica muy similar a la descrita en el apartado siguiente, los 50 ítems del cuestionario aplicado en el 2003 fueron revisados cuidadosamente. Fruto de este análisis, algunos ítems fueron modificados en mayor o menor grado y los que presentaban índices mediocres fueron eliminados. Éstos se muestran en gris en la columna penúltima de la tabla 1.

- *Segunda administración del cuestionario: año 2005.*

Mientras la administración del cuestionario en el 2003 tenía como principal objetivo la depuración y validación de los ítems, en 2005 se pone especial énfasis en la utilidad del cuestionario para la evaluación de la experiencia docente. En la nueva experiencia, por tanto, el cuestionario se pasa como pretest y postest. Otra importante diferencia entre ambos años reside en que los criterios para analizar la bondad de cada ítem en 2005 se hicieron más precisos, consensuados y cuantitativos que los adoptados en 2003. En este caso, se calcularon 22 índices, cuya descripción veremos a continuación.

En 2005, el cuestionario queda compuesto por 40 ítems (ver anexo 1) y se pasa a una muestra de 50 alumnos.

En el siguiente apartado se muestra con detalle cómo se llevó a cabo la revisión y validación de ítems. Ésta, creemos, es la aportación más importante de este tra-

bajo y supone el factor que da fuerza y consistencia a nuestra apuesta por este cuestionario como herramienta útil y eficaz para conocer las creencias de ciencias de los alumnos.

5. VALORACIÓN DE LOS ÍTEMS DE LA PRUEBA

Como se dijo anteriormente, una de las características metodológicas aplicadas para tratar de paliar las debilidades del COMVDS fue la de solicitar al entrevistado los siguientes tipos de respuestas:

- en el pretest, además de la elección de la respuesta correcta, una valoración de la percepción de la bondad de cada ítem.
- en el postest, además de la elección de la respuesta correcta, la reacción afectiva del entrevistado al conocer la respuesta dada por los expertos.

Esto dio lugar a un total de once índices en el pretest que son los siguientes:

6 índices son elaborados usando los datos de la primera respuesta a los ítems, cuando se busca la opción correcta:

- Ia. Frecuencia en % de elección de la opción a (respuesta correcta según orden de la base de datos). Se considera que en el rango $30 < Ia < 70$ el ítem no es ni muy fácil ni excesivamente difícil; su probabilidad de discriminar es mayor y puede ser candidato a buen ítem dependiendo de los demás parámetros.
- Ib. Frecuencia en % de elección de la opción b (primer distractor del ítem). Si cumple bien su función, es deseable que $Ib > 20$.
- Ic. Frecuencia en % de elección de la opción c (segundo distractor del ítem). Si cumple bien su función, es deseable que $Ic > 20$.

El modelo ideal es que haya reparto entre las 3 opciones (a, b y c) pero también es bastante aceptable que $Ib + Ic > 30$. Si un distractor absorbe más frecuencia que la respuesta correcta, es necesario hacer un análisis cualitativo del ítem por si realmente no estuviera bien redactado o, incluso, se hubiera deslizado algún error o sesgo.

- I0. Frecuencia en % de alumnos que optan por no responder. El modelo ideal es que sea 0 o lo más bajo posible. Este dato, combinado con algunos índices de valoración, puede detectar ítems mal contruidos.

ID. Índice de discriminación. Correlación de resultados del ítem con totales. Este índice se calcula considerando los aciertos de dos grupos de sujetos: del 27% con las mejores notas en el cuestionario (A) y del 27% de peores notas (B) y se aplica la siguiente fórmula $(A - B) / 0,27 N$, siendo $N = N.^{\circ}$ sujetos. Para disponer de una imagen más cualitativa y adecuada de lo

que significa este índice, hay que indicar que si todos los alumnos del grupo alto contestan correctamente al ítem y ninguno del grupo bajo, entonces $ID = 1$ (valor máximo de este indicador); si sucede lo contrario, $ID = -1$ (valor máximo negativo); si ambos grupos contestan por igual, $ID = 0$ (valor mínimo de discriminación). Se considera que $ID > 0,39$ es un indicador de un ítem excelente y con $ID > 0,20$, la discriminación comienza a ser aceptable (Ebel y Frisbie, 1986, citados por Backhoff, Larrazolo y Rosas, 2000).

IE. Índice de eficacia. Es también un índice de discriminación pero en vez de considerar el 27% del total sólo se consideran los sujetos del grupo superior e inferior que han acertado el ítem, por lo tanto la fórmula es: $IE = (A - B) / (A + B)$. Su tendencia es dar mayor discriminación que ID. A partir de $IE > 0,20$ se acepta una discriminación aceptable. El índice de eficacia es una medida del ID relativo al grupo de alumnos que han respondido, de modo que discrimina dentro del grupo más alto de la muestra y viene a matizar y complementar a ID.

5 índices son elaborados desde la segunda respuesta a los ítems, cuando se valora la percepción del entrevistado:

- Va. Porcentaje de alumnos que perciben el ítem bien construido. Cuanto más alto sea Va mejor percepción se tiene del ítem. Posiblemente la coherencia se perciba cuando el ítem esté expresado en términos adecuados al conocimiento del entrevistado.
- Vb. Porcentaje de alumnos que creen que les faltan opciones. Lo más probable es que esto ocurra cuando el entrevistado no percibe ninguna opción del ítem como correcta.
- Vc. Porcentaje de alumnos que creen que hay más de una respuesta correcta. Esto puede ser debido a un conocimiento poco preciso del tema que se aborda en el ítem o puede ser que, en efecto, sea así. Esta última opción se puede contrastar parcialmente mediante un análisis de los índices Ia, Ib, Ic e I0.
- Vd. Porcentaje de alumnos que no perciben el ítem bien planteado por motivos diferentes a los recogidos por Vb o Vc.

En cualquier caso, la suma de Vb, Vc y Vd debe ser baja si se quiere considerar el ítem como bueno, mientras Va debe ser alta.

Ve. Porcentaje de alumnos que no dan opinión sobre el ítem. Puede ser porque resulte indiferente o porque no se tiene mucho que decir sobre el contenido del mismo.

Con estos índices, el ítem idóneo sería aquel cuyas alternativas fueran elegidas por el mismo porcentaje de alumnos ($Ia = Ib = Ic$), que no hubiera sido dejado en blanco por ningún alumno ($I0 = 0$), y que tuviera el valor de 100 como índices de discriminación y de eficacia. Además, debería

haber sido percibido por todos los alumnos como un ítem bien construido (Va debería valer 100) y el resto de los índices relativos a la valoración del ítem deberían ser nulos.

Del mismo modo, se calcularon otros 11 índices ofrecidos en la administración del postest.

Los 6 primeros, los referentes a la elección de la respuesta correcta, son los mismos que en el pretest. Pero en el postest, los 5 siguientes no son fruto de solicitar una valoración del ítem sino que, dado que esta experiencia docente tenía también objetivos de aprendizaje, resultaban de solicitar al alumno su impresión sobre el ítem una vez que escuchaba la respuesta que eligen los expertos. Estos cinco nuevos índices son los siguientes:

- Sa. Porcentaje de alumnos cuya respuesta coincide con la de los expertos.
- Sb. Porcentaje de alumnos que al escuchar la respuesta de los expertos sienten que les gustaría mantener su opción aunque ésta no sea la misma.
- Sc. Porcentaje de alumnos que les parece mejor la respuesta de los expertos que la suya y por lo tanto prefieren cambiar su respuesta.
- Sd. Porcentaje de alumnos que al escuchar la respuesta de los expertos no se deciden por nada y generalmente se sienten confusos o sienten conflicto.
- Se. Porcentaje de alumnos que no responden a alguna de las anteriores opiniones. Es usual alguna de estas dos razones: a) no tienen una buena opinión del ítem o b) no quieren implicarse con él porque no tienen nada que decir.

Los anteriores índices sobre reacción afectiva al escuchar la respuesta que eligen los expertos, sobre todo expresan el grado de implicación afectiva del alumno para responder al ítem, que sería tanto mayor en la medida que Sb, Sc y Sd fueran altos y más bajo fuera Se. Su inserción en la experiencia docente obedece a actividades posteriores de enseñanza y aprendizaje ligadas a la VdC, pero se percibe enseguida que igualmente es una valiosa información que nos habla de la persuasión o atractivo del ítem para implicar al alumno en su aprendizaje.

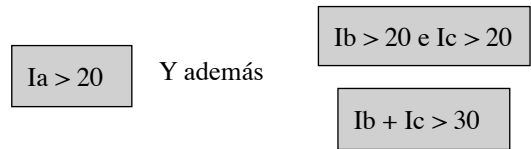
El problema con el que nos encontramos para valorar los ítems a partir de los índices calculados radica en la naturaleza diferente de estos índices, pues representan categorías numéricas incompatibles (unos reflejan decisiones cognoscitivas, otros opiniones y otras tendencias afectivas). ¿Qué se puede hacer entonces para valorar la bondad y eficacia de los ítems?

La mejor solución encontrada implica agruparlos en factores que representan aspectos diversos entre sí pero que, por distintos motivos, son importantes en la valoración de la bondad y eficacia del ítem. La suma ponderada de estos factores, teniendo en cuenta su importancia relativa en el cómputo global, puede aportarnos un índice cuantitativo de la calidad del ítem.

La fórmula alcanzada finalmente y que vamos a exponer a continuación fue obtenida por consenso entre jueces. En ella intervienen seis factores que se definen a continuación.

Factor 1: Está relacionado con la distribución de porcentajes de respuestas entre las diferentes opciones del ítem. Lo ideal de un ítem es que obtuviera un reparto equitativo entre sus alternativas, lo que matemáticamente se corresponde con $Ia = Ib = Ic = 33\%$. Esto también supone que $I0 = 0$, es decir, que nadie deja el ítem sin responder.

Los valores mínimos estimados para que este factor contabilice positivamente ($\frac{1}{2}$ o 1) en la valoración del ítem son:



Lo anterior asegura que simultáneamente se den dos condiciones: por un lado, que la respuesta correcta (a) obtenga un porcentaje mínimo y, por otro que las opciones que distraen (b y c) también lo tengan con el fin de asegurar que juegan bien su papel de ser plausibles para los menos expertos. Si estas condiciones se cumplen al límite de los valores señalados, el factor adquiere el valor $\frac{1}{2}$; si se cumplen con holgura, el valor 1; y 0 si no se cumplen.

Factor 2: Es el factor que recoge la capacidad de discriminar que tiene cada ítem. Viene determinado por los índices ID e IE, los cuales recogen la capacidad de discriminación de cada ítem particular en relación con los resultados globales de la prueba. Lo ideal es que estos índices tomen el valor máximo de 100, pero lo mínimo aceptable para que un ítem sea bueno es que su índice de discriminación sea mayor de 20.

En ítems donde $Ia < 20$ (ítem difícil), se le da más peso a IE que a ID dado que IE sólo considera la submuestra de los que responden correctamente al ítem. Además, si $Ia < 20$ e $IE > 20$, revela que el ítem, aunque difícil, tiene buena capacidad discriminatoria y puede jugar un papel importante para evaluar el incremento de la por aprendizaje cuando forme parte del postest. Por tanto, este factor vale:

Factor 2	0	0,5	1
Si Ia > 20	ID < 20 e IE < 20	20 < ID < 30 20 < IE < 30	ID > 30 e IE > 30
Si Ia < 20	IE < 20	20 < IE < 30	IE > 30

Factor 3: Este índice hace referencia a la valoración que los alumnos hacen del ítem y de su construcción. Influyen en su tasación los índices de valoración Va, Vb,

Vc, Vd y Ve e I0. Este último representa el porcentaje de sujetos que deja el ítem sin responder, y es una aportación importante para este factor por ser precisamente la percepción deficiente de un ítem la razón más frecuente para no responder al mismo.

Las referencias cuantitativas que permitieron una valoración convergente a los grupos de trabajo fueron las siguientes:

Factor 3	0	0,5	1
	$Va < 40$	$50 > Va > 40$	$50 < Va$
	$Vb + Vc + Vd > 30$	$20 < Vb + Vc + Vd < 30$	$20 > Vb + Vc + Vd$
	$I0 > 30$	$20 < I0 < 30$	$20 > I0$

Factor 4: Es el mismo factor 1 aplicado a los índices del postest.

Factor 5: Es el mismo factor 2 aplicado a los índices del postest.

Factor 6: Revela el grado de implicación afectiva del alumno para responder al ítem. Para ello se considera, por un lado, *Se* que recoge la indiferencia ante el ítem (deberá ser bajo) y, por otro, la suma de *Sc* y *Sd* que recogen respectivamente el deseo de cambiar la opción previamente elegida o la indecisión ante el conflicto ocasionado al escuchar qué opción cogerían los expertos.

Los valores de referencia acordados para este factor son:

Factor 6	0	0,5	1
	$Se > 20$	$20 < Se < 10$	$Se < 10$
	$Sc + Sd < 30$	$40 > Sc + Sd > 30$	$Sc + Sd > 40$

Valoración global de la bondad: La valoración numérica del ítem se alcanza sumando los valores que toma cada factor. Es un número comprendido entre 0 y 6. La cifra obtenida se interpreta de la siguiente manera:

Si está en el intervalo [0; 2,5], nivel inferior de bondad del ítem

Si está en el intervalo [2,6; 4,5], nivel medio de bondad del ítem

Si está en el intervalo [4,6; 6], nivel superior de bondad.

Se considera que tres niveles son suficientes para los objetivos de la investigación (nótese que, finalmente, se considera nivel 1 al ítem de más calidad y nivel 3 al de menos calidad):

NIVEL DE CALIDAD DEL ÍTEM		
Nivel 1	Nivel 2	Nivel 3
[4,6 - 6]	[2,6 - 4,5]	[0 - 2,5]

Para aumentar la objetividad del proceso de valoración de la calidad de los ítems, éste fue efectuado por cuatro grupos de futuros profesores de secundaria y dos expertos. En la tabla 2 se sintetiza el proceso de valoración y clasificación de los ítems. En ella, se recogen los índices de calidad adjudicados a cada uno de los 40 ítems del cuestionario por los cuatro grupos de trabajo (ver respectivamente las columnas 3, 4, 5 y 6) y por los dos expertos (columna 8). De la columna 1 a 3 se ha colocado respectivamente el orden de cada ítem según aparece en la prueba, su código único respecto a la base de datos y su tipo de estructura. En la columna 9 se ha colocado la media aritmética y, finalmente en la última, el nivel de calidad asignado aplicando el criterio dado en la tabla anterior.

Conviene destacar que, en este proceso, las medidas adoptadas para hacer poco probable una clasificación incorrecta de los ítems han sido las siguientes:

- Los criterios cuantitativos establecidos llevan a valorar los ítems de un modo bastante convergente, como se puede percibir claramente en las columnas 4 - 8 de la tabla 2.
- Puesto que diferenciar tres niveles de calidad es suficiente para lograr los objetivos de la investigación, la escala original para valorar cada ítem se ha transformado de [0-6], mediante intervalos, a una escala [1-3].

La combinación de ambas medidas hace que, por ejemplo, un ítem colocado en el nivel 3 sea poco probable que con otro sistema de jueces pudiera estar en el nivel 2.

En la tabla 2 se puede apreciar que hay 13 ítems de nivel 3, y por tanto, de baja calidad. Lo más correcto sería excluirlos de posibles análisis posteriores, por ejemplo, para calcular las diferencias entre pretest y postest, evitando enmascarar los diferentes progresos cognitivos. Estos análisis deberían realizarse sólo con los ítems 1 y 2 que se han mostrado más válidos y discriminatorios.

En la tabla 3 se muestran los códigos de los ítems clasificados por niveles, así como unos ejemplos en los que se ha procurado elegir ítems representativos de cada nivel. Con los comentarios se ha pretendido simular el proceso seguido en la selección de los ítems que conforman el cuestionario final.

Finalmente, otro resultado destacado sobre la calidad de los ítems, en este caso sobre su fiabilidad, es el que se expone en la tabla 4. Se muestran las correlaciones entre los índices de las pruebas del 2003 y 2005, y las correlaciones entre los índices del pretest y del postest del 2005.

Las correlaciones indican el grado de estabilidad o fiabilidad de los ítems para dar índices semejantes ante diferentes muestras o momentos del aprendizaje. Correlaciones razonablemente altas permitirán abrazar la esperanza de que poco a poco la base de datos se podrá ir llenando de ítems buenos o muy buenos. Este análisis es especialmente importante puesto que una correlación baja mostraría que el proceso de clasificación de ítems en tres niveles no tendría mucho sentido, pues para otra muestra puede cambiar drásticamente.

Como se puede observar en la tabla 4, todas las correlaciones entre índices semejantes, salvo pocas excepciones que comentaremos, son significativas al menos al 95% de confianza. Los datos más significativos que muestra la tabla 4 son:

- Observando los índices Ia, Ib, Ic e I0, vemos que las correlaciones son buenas o muy buenas (salvo Ib entre 2003 y 2005). Las elevadas correlaciones entre el pretest y postest del 2005 muestran un comportamiento uniforme de la prueba cuando es pasada en dos momentos diferentes del proceso de aprendizaje, y, por tanto, son indicativas de una alta fiabilidad. Debido a que entre las pruebas de 2003 y 2005 se retocaron algunos ítems y otros se eliminaron, las correlaciones no son tan altas aunque siguen siendo significativas.

- Resultan sorprendentes las buenas correlaciones de los índices V (percepción de la bondad de la construcción del ítem) y S (reacción afectiva al conocer la respuesta correcta), pues tanto unos como otros recogen percepciones subjetivas del entrevistado. Salvo la falta de correlación del índice Sc, número de sujetos que tras escuchar la respuesta correcta desean cambiar la suya propia, los demás índices dan una visión fiable de los ítems, corroborando la imagen dada por los índices I que por ser más directos son más importantes para expresar la fiabilidad.

- La baja correlación entre índices de discriminación ID es explicable si tenemos en cuenta que este índice se calcula por la correlación entre los resultados de un ítem y el rendimiento global de la prueba. Este último es sensible bien a modificaciones parciales de la prueba, bien a procesos de aprendizaje. Así pues, lo que realmente cambia en el índice ID entre 2003 y 2005 o entre el pretest y postest del 2005 es la discriminación de la prueba y no tanto la de muchos de los ítems que la conforman (bien mostrada con Ia, Ib, Ic, I0).

6. CONCLUSIONES SOBRE EL CUESTIONARIO COMVDC

La experiencia aplicando el COMVDC en actividades docentes de enseñanza y evaluación del aprendizaje desde 2003 a 2005, ha aportado una buena cantidad de datos, suficientes para ofrecer conclusiones significativas ligadas al comportamiento de este cuestionario y, sobre todo, de sus posibilidades futuras.

- La característica más destacada del proceso seguido es la cualidad que se deriva de usar mecanismos de autocorrección. Esta cualidad se concreta en la capacidad para discriminar ítems de baja calidad, frente a otros de media y alta (ver última columna de la tabla 2 para los ítems del cuestionario del anexo 1). Esto es importante para conocer las posibilidades del instrumento de medida que se está usando y, sobre todo, mejorarlo.

Lo anterior ha sido posible gracias a la técnica de administrar COMVDC solicitando, en el pretest, dos respuestas para cada ítem: una que pretende encontrar la respuesta correcta, y otra valorando la calidad del ítem, y, en el caso del postest, valorando la reacción afectiva al conocer la respuesta correcta; y calculando unos indicadores que perfilan una imagen sobre su comportamiento.

Lo más importante es que esta imagen muestra cierta estabilidad cuando se administra en sucesivas ocasiones (ver tabla 4), lo cual muestra, por un lado, el grado de significación de los índices que perfilan el comportamiento de cada ítem y, por otro, abre la posibilidad nada desdeñable de disponer de una base de datos compuesta sólo de ítems de buena calidad.

- En general, el COMVDC y la mecánica de autoevaluación comentada componen un sistema que converge hacia un cuestionario cada vez más válido y fiable.

Los elementos básicos de esta mecánica de autoevaluación han sido los siguientes:

- a) partir de la sistemática de contextos que se pretende evaluar.
- b) diseñar ítems ligados a cada contenido y a una estructura de opciones determinada.
- c) administrar el cuestionario solicitando respuestas de conocimiento del contenido (tipo I), de valoración del ítem (tipo V) y de afectividad ante el conocimiento experto (tipo S).
- d) calcular el índice de calidad de cada ítem y proceder a la depuración de los mismos, lo que significa eliminar unos ítems, retocar otros y validar el resto.
- e) alimentar la base de ítems con otros nuevos (si es posible, consensuados por un grupo de expertos), siempre ligados a los contenidos que interesa evaluar.
- f) Anotar en la base de ítems el historial de índices y comportamientos adoptado en sus sucesivas aplicaciones.

Esto conlleva un cambio de actitud para el constructor de pruebas de opciones múltiples. Cada ítem se debe ver como una propuesta que se hace a modo de tentativa hipotética, incluso aunque se perciba inicialmente como muy bueno. Siempre habrá que esperar a confirmarlo tras su aplicación. Esto hace que más importante que construir nuevos ítems será disponer de una mecánica de

falsación para evaluarlos, confiando más en ella que en la supuesta habilidad para diseñar los ítems.

– El COMVDC tiene muy buenas cualidades para ser aplicado en contextos docentes, tales como la posibilidad de ser administrado en una hora de clase y de poderse integrar en el proceso de enseñanza como una herramienta didáctica eficaz.

Concretamente, los resultados del COMVDC sobre cada alumno permiten identificar no sólo los progresos de aprendizaje (al comparar los resultados entre pretest y postest), sino que además, al estar las opciones de algunos ítems estructuradas según tendencias epistemológicas, se pueden hacer, por ejemplo, agrupamientos de alumnos de tendencia constructivista o hacer un seguimiento individual de tendencias epistemológicas. Especificando más, algunas posibilidades que el profesor-investigador tiene con el uso de este cuestionario son:

– Identificar los progresos del aprendizaje experimentados tras la enseñanza sobre la naturaleza de la ciencia, administrando el cuestionario antes y después del proceso de enseñanza. Con esta finalidad, se deberían seleccionar los mejores ítems del cuestionario (que tienen niveles de calidad 1 y 2, según la última columna de la tabla 2).

– Seleccionar los ítems de mejor calidad con estructura constructivista-racionalista-empirista (ítems con estructura C3, según la tercera columna de la tabla 2) y calcular los rendimientos globales de la muestra para este parcial de ítems. Dado que la opción más racionalista (b) suele mostrar índices significativamente más bajos que las opciones (a) y (c), es fácil realizar agrupamientos de alumnos más constructivistas y más empiristas y, después, analizar su evolución con los resultados en el postest.

– Se puede hacer también la misma operación anterior para los ítems con estructura organicista-reduccionista (ítems con estructura O, según la tercera columna de la tabla 2)

– Del mismo modo, se pueden analizar las correlaciones entre tendencias constructivistas y organicistas o comprobar hasta qué punto estas visiones son estructuras estables en la mente del alumno, analizando en estudios horizontales su comportamiento más o menos constante para los diferentes contextos en los que se manifiesta la VdC.

Aunque algunos de estos análisis se han llevado a cabo, las limitaciones en la extensión del artículo impiden mostrar los resultados.

REFERENCIAS BIBLIOGRÁFICAS

- ABAD, F.J., ATENCIA, J., GARCÍA, C., HONTANGAS, P., OLEA, PONSODA, V., REVUELTA, J., SUERO, M. y XI-MÉNEZ, C. (2006). Proyecto de innovación docente. Ayuda a la creación de exámenes. Disponible en <<http://www.uam.es/docencia/ace/>>.
- ACEVEDO, J.A. (1992). Cuestiones de sociología y epistemología de la ciencia. La opinión de los estudiantes. *Revista de Educación de la Universidad de Granada*, 6, pp. 167-182.
- ACEVEDO, J.A. (1994). «Los futuros profesores de enseñanza secundaria ante la sociología y la epistemología de las ciencias», en: *Revista Interuniversitaria de Formación del Profesorado*, 19, pp. 111-125. Versión electrónica corregida y actualizada en Sala de Lecturas CTS+I de la OEI <<http://www.campusoei.org/salactsi/acevedo8.htm>>, 2001.
- ACEVEDO, J.A., ACEVEDO, P., MANASSERO, M.A. y VÁZQUEZ, A. (2001). «Avances metodológicos en la investigación sobre evaluación de actitudes y creencias CTS», en: *Revista Iberoamericana de Educación*, edición electrónica De los Lectores (4-6-2001) <<http://www.campus-oei.org/revista/deloslectores/Acevedo.PDF>>.
- AIKENHEAD, G.S. (1988). An analysis of four ways of assessing student beliefs about STS topics. *Journal of Research in Science Teaching*, 25(8), pp. 607-629.
- AIKENHEAD, G.S., FLEMING, R.W. y RYAN, A.G. (1987). High-school graduates' beliefs about sciencetechnology-society. I. Methods and issues in monitoring student views. *Science Education*, 71(2), pp. 145-161.
- AIKENHEAD, G.S. y RYAN, A.G. (1992). The development of a new instrument: 'Views on Science-Technology-Society' (VOSTS). *Science Education*, 76(5), pp. 477-491.
- BACKHOFF, E. LARRAZOLO, N. y ROSAS, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA), *Revista Electrónica de Investigación Educativa*, 2(1), Consultado el 15 de julio de 2007 en: <<http://redie.uabc.mx/vol2no1/contenido-backhoff.html>>.
- FERNÁNDEZ, I., GIL, D., CARRASCOSA, J., CACHAPUZ, A. y PRAIA, J. (2002). Visiones deformadas de la ciencia transmitidas por la enseñanza. *Enseñanza de las Ciencias*, 20(3), pp. 477-488.
- IZQUIERDO, M. (2000). Fundamentos Epistemológicos. pp. 35-64. En F.J. Perales y P. Cañal. (eds.). *Didáctica de las Ciencias Experimentales*. Alcoy: Marfil.
- LAFOURCADE, P.D. (1972). *Evaluación de los aprendizajes*. Madrid: Cincel.

- LEDERMAN, N.G., ABD-EL-KHALICK F., BELL, R.L. y SCHWARTZ, R.S. (2002). Views of Nature of Science Questionnaire: Towards valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), pp. 551-581.
- LEDERMAN, N. G., WADE, P. D. y BELL, R. L. (1998). Assessing the nature of science: What is the nature of our assessments? *Science and Education*, 7 (6), pp. 595-615.
- MANASSERO, M. A. y VÁZQUEZ, A. (2002). Instrumentos y métodos para la evaluación de las actitudes relacionadas con la Ciencia, la Tecnología y la Sociedad. *Enseñanza de las Ciencias*, 20(1), pp. 15-27.
- MARÍN, N. (2003). Visión constructivista dinámica para la enseñanza de las ciencias. *Enseñanza de las Ciencias*, número extra, pp. 43-55.
- MARÍN, N. (2005). *La enseñanza de las ciencias en Educación Infantil*. Grupo Editorial Universitario: Granada.
- McCOMAS W.F., CLOUGH, M.P. y ALMAZROA, H. (1998). *The Role And Character of The Nature of Science in Science Education*, pp. 3-39, en W.F. McComas (eds.). *The Nature Of Science In Science Education. Rationales and Strategies*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- OSBORNE, J., COLLINS, S., RATCLIFFE, M., MILLAR, R. y DUSCHL, R. (2003). What «Ideas-about-Science» Should Be Taught in School Science? A Delphi Study of the Expert Community. *Journal of Research in Science Teaching*, 40, pp. 692-720.
- RODRÍGUEZ DIÉGUEZ, J.L. (1980). *Didáctica General*, «1. Objetivos y evaluación». Madrid: Cincel-Kapelusz.
- RUBBA, P.A. y HARKNESS, W.L. (1993). Examination of preservice and in-service secondary science teachers' beliefs about Science-Technology-Society interactions. *Science Education*, 77, pp. 407-431.
- SCHONEWEG-BRADFORD, C.S., RUBBA, P.A. y HARKNESS, W.L. (1995). Views about Science-Technology-Society interactions held by college students in General Education Physics and STS courses. *Science Education*, 79(4), pp. 355-373.
- VÁZQUEZ, A., ACEVEDO, J.A. y MANASSERO, M.A. (2000). Progresos en la evaluación de actitudes relacionadas con la Ciencia mediante el cuestionario de opiniones CTS. Consultado el 21 de noviembre de 2006 en: <<http://www.oei.es/salactsi/acevedo6.htm>>.
- VÁZQUEZ, A., ACEVEDO, J.A. y MANASSERO, M.A. (2004). Consensos sobre la naturaleza de la ciencia: evidencias e implicaciones para su enseñanza. *Revista Iberoamericana de Educación*, edición digital en <<http://www.campus-oei.org/revista/deloslectores/702Vazquez.PDF>>.
- VÁZQUEZ, A. y MANASSERO, M. A. (1997). Una evaluación de las actitudes relacionadas con la ciencia. *Enseñanza de las Ciencias*, 15 (2), pp. 199-213.
- VÁZQUEZ, A. y MANASSERO, M.A. (1998). *Opinions sobre ciència, tecnologia i societat*. Palma de Mallorca: Conselleria d'Educació, Cultura i Esports.
- VÁZQUEZ, A. y MANASSERO, M.A. (1999). Response and scoring models for the 'Views on Science-Technology-Society' instrument. *International Journal of Science Education*, 21(3), pp. 231-247.
- VÁZQUEZ ALONSO, A., MANASSERO, M. A. y ACEVEDO, J. A. (2005). Análisis cuantitativo de ítems complejos de opción múltiple en ciencia, tecnología y sociedad: Escalamiento de ítems. *Revista Electrónica de Investigación Educativa*, 7 (1). Consultado el 22 de noviembre de 2006 en: <<http://redie.uabc.mx/vol7no1/contenido-vazquez.html>>.

[Artículo recibido en mayo de 2008 y aceptado en julio de 2008]

Tabla 1: Sistemática de contextos para el cuestionario sobre VdC

Categoría	Visión +	Visión -	Códigos	n.º
1	Contexto donde surge y se aplica la ciencia (interacciones CTS)			9
1a	Relaciones CTS	La ciencia es un conocimiento comprometido	7111 (On), 7112 (On) 7113 (On)	3
1b	Relaciones entre ciencia (C) y sociedad (S)	La ciencia surge de los problemas del momento	7114 (On), 7115 (On)	2
1c	Relación entre tecnología (T) y ciencia	Existen fuertes vínculos entre T-C	7116 (On), 7117 (On)	2
1d	¿Quién gestiona y aplica este conocimiento?	Existen compromisos e intereses	7118 (On), 7119 (On)	2
2	Fase privada (actividad en la fase de descubrimiento)			9
2a	¿Afectan al científico creencias culturales (morales, religiosas) de su región?	Al científico le afectan los compromisos, creencias e intereses que hacen que su actividad no sea exclusivamente racional.	7121 (Os), 7122 (Os)	2
2b	¿Afectan problemas sociales y políticos al científico?	Visión comprometida con el contexto social, cultural y económico, incluso personal	7123 (Os), 7124 (Os)	2
2c	¿Influyen en el científico su entorno cotidiano e intereses personales?		7125 (Os), 7126 (Os) 7127 (Os)	3
2d	¿Siguen los científicos patrones metodológicos establecidos?	También se da el azar, ensayos, vuelta atrás, rectificaciones...	7228 (Or), 7129 (Or)	2
3	Interacción entre fase privada y pública (formación y difusión)			9
3a	Sobre el experto de la fase privada que hace aportaciones en la pública	Buena formación e integración en el grupo. La comunidad científica regula la producción	7131 (Oi), 7132 (Oi)	2
3b	Relaciones entre el trabajo publicado y el realizado	Procesos complejos de asimilación orgánica no regidos en exclusiva por lo racional	7133 (Oa), 7134 (Oa)	2
3c	¿Cómo es la mecánica que regula las incorporaciones privadas a la ciencia pública?	Complejo proceso de asimilación de las aportaciones individuales	7135 (Oi), 7136 (Oi)	2
3d	Secuencia para integrar las aportaciones privadas en lo público	Unas veces se debe a asimilaciones lineales, otras a saltos revolucionarios	7137 (Oa), 7138 (Oa) 7139 (Oa)	3
4	Fase de justificación (la ciencia como producto)			27
4a	¿Es diferente a otros conocimientos? ¿A qué realidad se refiere?	Es el conocimiento más contrastado (C). La realidad contruida cambia cuando cambian las teorías (O)	7211 (C3), 7212 (C3) 7213 (C3), 7314 (C3) 7215 (O3), 7216 (O3) 7217 (O3), 7218 (O3) 7219 (O3)	9
4b	¿Cómo se pueden explicar los éxitos de las teorías de ciencias?	Constante confrontación entre teoría y datos (C)	7221 (C3), 7222 (C3) 7223 (C3), 7324 (C3) 7225 (Ce), 7226 (Ce) 7227 (Ce), 7228 (Ce) 7229 (Ce)	9
4c	La ciencia ¿descubre o inventa? ¿refleja o interpreta la realidad?	Los datos no son más creíbles que la teoría	7231 (Ce), 7232 (Ce) 7233 (Ce), 7234 (Ce)	4
4d	¿Qué dinámica se sigue para aceptar o refutar las ideas y teorías científicas? (T1 > T2)	Coherencia entre partes (teóricas y empíricas). Análisis y síntesis	7235 (O3), 7236 (O3) 7237 (O3), 7238 (O3) 7239 (O3)	5

Tabla 2: Valoración de ítems sobre VdC

Orden en prueba	Código ítem	Tipo de ítems	Grupo n.º 1	Grupo n.º 2	Grupo n.º 3	Grupo n.º 4	Grupo n.º 5	Media aritmética	Nivel de calidad
n.º 01	7111	On	1	1	1	1	2	1,2	3
n.º 02	7113	On	6	6	5	6	6	5,8	1
n.º 03	7114	On	1	1	1	1	2,5	1,3	3
n.º 04	7118	On	3,5	3	3	3	4,5	3,4	2
n.º 05	7119	On	3,5	2	2	3	4	2,9	2
n.º 06	7121	Os	4	3	3	3	5	3,6	2
n.º 07	7122	Os	5	4	4	4	5,5	4,5	2
n.º 08	7124	Os	4	3	3	3	4	3,4	2
n.º 09	7125	Os	2	2	2	2	2	2	3
n.º 10	7128	Or	1,5	1,5	1	1	2	1,4	3
n.º 11	7129	Or	4	3,5	4	4	5,5	4,2	2
n.º 12	7131	Oi	3,5	2	2	2	4	2,7	3
n.º 13	7132	Oi	3	1,5	2	2	3	2,3	3
n.º 14	7133	Oa	3,5	2,5	2	3	4	3	2
n.º 15	7136	Oi	1	1	1	1	1,5	1,1	3
n.º 16	7137	Oa	5	5	4,5	5	5	4,9	1
n.º 17	7211	C3	3,5	2,5	3	3	2,5	2,9	2
n.º 18	7212	C3	4	3,5	3	3,5	4,5	3,7	2
n.º 19	7213	C3	5	5,5	5	5	5,5	5,2	1
n.º 20	7214	C3	2	1	1	2	2	1,6	3
n.º 21	7215	O3	2	1,5	1	2	3	1,9	3
n.º 22	7216	O3	4	4	4	4	4,5	4,1	2
n.º 23	7217	O3	4,5	4	4	4	5	4,3	2
n.º 24	7218	O3	1,5	2	2	2	2,5	2	3
n.º 25	7221	C3	4	3,5	3	3,5	4,5	3,7	2
n.º 26	7223	C3	5	5	5	5	5,5	5,1	1
n.º 27	7224	C3	2	2	2	2	2,5	2,1	3
n.º 28	7225	Ce	6	6	5	5	6	5,6	1
n.º 29	7226	Ce	2	2,5	2	2	3,5	2,4	3
n.º 30	7227	Ce	3	3	3	3	3	3	2
n.º 31	7228	Ce	3,5	3,5	3	4	4	3,6	2
n.º 32	7229	Ce	3	3	3	3	3,5	3,1	2
n.º 33	7231	Ce	5	4,5	4	5	5	4,7	1
n.º 34	7233	Ce	4,5	5,5	4	5	6	5	1
n.º 35	7234	Ce	5	4,5	4	5	6	4,9	1
n.º 36	7235	O3	3,5	3,5	3	3,5	4,5	3,6	2
n.º 37	7236	O3	4,5	4	3	4	5	4,1	2
n.º 38	7237	O3	2,5	2	2	2	1,5	2	3
n.º 39	7238	O3	3,5	3	3	3	3,5	3,2	2
n.º 40	7239	O3	4	4	3	4	5,5	4,1	2

Tabla 3: Clasificación y ejemplos de ítems de la prueba COMVDC

Nivel	Tipo	Códigos	Ejemplos
1	C (6)	n.º 19 7213 C3 n.º 26 7223 C3 n.º 28 7225 Ce n.º 33 7231 Ce n.º 34 7233 Ce n.º 35 7234 Ce	Comentario: Un buen ítem, algo difícil, pero discrimina muy bien y mantiene un comportamiento bastante fiable. Resultados en 2003: Ia:16 b:04 c:48 0:32 D:14 E:33 / Va:16 Vb:00 Vc:24 Vd:08 V0:52 / Sa:34 Sb:08 Sc:30 Sd:04 S0:21. Retocado para 2005, continúa su excelente rendimiento (Ia:20 b:16 c:52 0:09 D:33 E:71 / Va:67 b:05 c:14 d:04 e:10 / Ia:26 b:14 c:56 0:04 D:35 E:45 / Sa:24 b:36 c:33 d:05 e:05). Se percibe bien construido y tiene buena actitud para implicar al alumno en el aprendizaje. [C3] Dicen que «la experiencia es la madre de la ciencia» ¿es cierto? a) Sí. Teoría y datos siempre se están enfrentando b) No. Más bien es un esfuerzo racional coherente y consensuado c) Sí. La ciencia se construye desde la experimentación
	O (2)	n.º 02 7113 On n.º 16 7137 Oa	Comentario: Excelente ítem que une a su brevedad buena fiabilidad y discriminación. Resultados en 2003: Ia:56 b:16 c:16 0:12 D:28 E:33 / Va:44 Vb:04 Vc:08 Vd:00 V0:44 BB / Sa:60 Sb:13 Sc:08 Sd:13 Se:05. El ítem se retoca un poco para el 2005 y continúa dando buenos resultados (Ia:28 b:24 c:22 0:24 D:20 E:42 / Va:62 b:20 c:04 d:08 e:06 / Ia:30 b:30 c:20 0:20 D:28 E:66 / Sa:30 b:34 c:25 d:03 e:11). Cabe destacar la buena percepción que da al entrevistado así como su capacidad para implicarlo en el aprendizaje. [On] Un país ¿es más rico porque es puntero en ciencia o es puntero porque es más rico? a) Más bien lo segundo b) Más bien lo primero c) La relación riqueza-ciencia no es tan directa
2	C (6)	n.º 17 7211 C3 n.º 18 7212 C3 n.º 25 7221 C3 n.º 30 7227 Ce n.º 31 7228 Ce n.º 32 7229 Ce	Comentario: El ítem se elige en 2003 por su buen rendimiento (Ia:76 b:04 c:04 0:16 D:57 E:50 / Va:44 Vb:12 Vc:00 Vd:04 V0:40 Mb / Sa:73 Sb:04 Sc:21 Sd:00 Se:00). Retocado para 2005, el ítem sigue dando un perfil estable (Ia:75 b:00 c:11 0:13 D:40 E:30 / Va:73 b:03 c:04 d:12 e:08 / Ia:90 b:02 c:04 0:04 D:28 E:16 / Sa:92 b:00 c:07 d:00 e:03) y excelente discriminación. Aunque es bien percibido, baja al nivel 2 porque su índice Ia es muy alto y no distribuye bien entre las opciones. [Ce] Será cierta la frase «mayor o menor, siempre se comete error al medir» a) Sí, no hay manera de medir exacto b) Es falsa a nivel microscópico c) No, si se usan aparatos de precisión
	O (13)	n.º 04 7118 On n.º 05 7119 On n.º 06 7121 Os n.º 07 7122 Os n.º 08 7124 Os n.º 11 7129 Or n.º 14 7133 Oa n.º 22 7216 O3 n.º 23 7217 O3 n.º 36 7235 O3 n.º 37 7236 O3 n.º 39 7238 O3 n.º 40 7239 O3	Comentario: El comportamiento aceptable en 2003 (Ia:56 b:00 c:16 0:28 D:28 E:33 / Va:16 Vb:04 Vc:12 Vd:00 V0:68 / Sa:65 Sb:04 Sc:21 Sd:04 Se:05) le hace pasar al 2005 con algunos retoques donde ofrece resultados semejantes pero con una mejora sustancial de la discriminación (Ia:54 b:11 c:07 0:26 D:53 E:57 / Va:45 b:09 c:21 d:14 e:12 / Ia:68 b:12 c:06 0:12 D:64 E:52 / Sa:72 b:04 c:17 d:05 e:05). Su valoración baja, dado que su construcción no es bien percibida y no implica suficiente al alumno en su aprendizaje. [On] ¿Avanzaría más rápido la ciencia si estuviese controlada por las empresas privadas? a) Es posible, pero probable que se fomente aspectos de sus intereses b) Sí, al fomentar la competitividad avanza rápido en todas direcciones c) No, pues sólo desde la neutralidad se desarrolla la ciencia
3	C (3)	n.º 20 7214 C3 n.º 27 7224 C3 n.º 29 7226 Ce	Comentario: Se elige en 2003 por su rendimiento aceptable (Ia:55 b:10 c:20 0:13 D:22 E:45 / Va:79 b:00 c:01 d:08 e:11 / Sa:76 b:06 c:10 d:05 e:01). Retocado para 2005, no ofrece los resultados que se esperaban (Ia:60 b:05 c:22 0:11 D:06 E:05 / Va:69 b:00 c:06 d:06 e:19 / Ia:84 b:02 c:14 0:00 D:21 E:13 / Sa:82 b:00 c:13 d:05 e:03). Baja capacidad de discriminación y distribución. [Ce] Entre teoría y datos experimentales, ¿a qué da más credibilidad la ciencia? a) Ambos tienen semejante credibilidad b) A la teoría, síntesis de muchos datos c) A los datos, fiel reflejo de la naturaleza
	O (10)	n.º 01 7111 On n.º 03 7114 On n.º 09 7125 Os n.º 10 7128 Or n.º 12 7131 Oi n.º 13 7132 Oi n.º 15 7136 Oi n.º 21 7215 O3 n.º 24 7218 O3 n.º 38 7237 O3	Comentario: Ítem difícil. Elegido para 2005 por su alto IE (Ia:16 b:32 c:00 0:52 D:14 E:99 / Va:04 Vb:16 Vc:32 Vd:00 V0:48 BB / Sa:30 Sb:17 Sc:13 Sd:34 Se:04) esperando que al ser retocado cambie y mejore su comportamiento, algo que, efectivamente, ocurre en 2005 (Ia:73 b:00 c:11 0:15 D:20 E:14 / Va:69 b:01 c:17 d:04 e:08 / Ia:86 b:00 c:10 0:04 D:14 E:08 / Sa:88 b:04 c:07 d:00 e:03). Se muestra como ítem fácil pero su baja capacidad para discriminar, distribuir e implicar lo lleva al nivel 3. [O3] De pronto, un experimento aporta datos que contradicen la actual teoría de la luz. ¿Qué es lo más probable que revisen en un primer momento los científicos? a) La fiabilidad de los datos de dicho experimento b) Aspectos de la teoría pero sin tocar su núcleo principal c) La misma teoría para ver en qué ha podido fallar

Tabla 4: Correlaciones de índices de COMVDC

<i>Correlaciones entre 2003 y 2005</i>				
Ia03-Ia05	Ib03-Ib05	Ic03-Ic05	Io03-Io05	ID03-ID05
0,586**	0,25	0,466**	0,429**	0,198
Va03-Va05	Vb03-Ib05	Vc03-Vc05	Vd03-Vd05	Ve03-Ve05
0,510**	0,406**	0,547*	0,378*	0,488**
Sa03-Sa05	Sb03-Sb05	Sc03-Sc05	Sd03-Sd05	Se03-Se05
0,559**	0,307	-0,044	0,363*	0,505*
<i>Correlaciones entre pretest y posttest del 2005</i>				
Ia-Ia'	Ib-Ib'	Ic-Ic'	Io-Io'	ID-ID'
0,939**	0,909**	0,858**	0,766**	0,303

Cuestionario sobre creencias de ciencias (VdC)		Fecha	31-01-2005
Apellidos			Nombre

1. La principal razón por la que un país debería invertir en ciencia es:
 - a) Por el valor cultural del conocimiento de ciencias
 - b) Por su utilidad, ya que permite mayor control y beneficio del medio
 - c) Para satisfacer el impulso de conocer lo desconocido
2. Un país ¿es más rico porque es puntero en ciencia o es puntero porque es más rico?
 - a) Más bien lo primero
 - b) La relación riqueza-ciencia no es tan directa
 - c) Más bien lo segundo
3. ¿Con qué intensidad se relaciona la ciencia (C) y la tecnología (T)?
 - a) Media, T es orientada con lo que C va descubriendo
 - b) Débil, sus fines e intereses son diferentes
 - c) Fuerte, los progresos de cada una ayudan a la otra
4. ¿Avanzaría más rápido y mejor la ciencia si estuviese controlada por las empresas privadas?
 - a) No, pues sólo desde la neutralidad se desarrolla la ciencia
 - b) Es posible, pero se fomentarán más temas de sus intereses
 - c) Sí, al fomentar la competitividad avanza rápido en todas direcciones
5. ¿Son responsables los científicos del daño que pueda resultar de sus descubrimientos?
 - a) A veces, cuando conocen sus aplicaciones
 - b) Sí, pues deben evitar que puedan producir daños
 - c) No, sólo de que sus descubrimientos sean certeros
6. ¿Cuál es la razón principal por la que el científico suele actuar con honradez?
 - a) De ese modo va a tener el reconocimiento de los demás
 - b) Para poder crear un conocimiento verdadero
 - c) Sabe que sus resultados van a ser comprobados por otros
7. ¿Influye en los descubrimientos del científico sus creencias religiosas?
 - a) En parte sí, depende del arraigo de tales creencias
 - b) No, la razón científica no depende de creencias religiosas
 - c) No, éstos se basan en hechos y en experiencias, no en creencias
8. El refrán «la paciencia es la madre de la ciencia» hace referencia a que los expertos, cuando hacen ciencia:
 - a) Confrontan sus teorías con paciente perseverancia
 - b) Afrontan los problemas con paciencia racional sin nervios
 - c) Siguen con paciencia un plan metódico aunque los datos sean contrarios
9. La formación y el carácter personal del científico ¿puede influir en sus descubrimientos?
 - a) No, pues los descubrimientos están basados en hechos
 - b) Sí, expertos con igual formación no actúan o inventan lo mismo
 - c) No, la verdad científica es independiente del científico
10. Investigando según un plan de trabajo ¿qué suelen hacer los científicos cuando aparecen problemas no previstos?
 - a) Modifican el plan dando prioridad a la teoría sobre los datos
 - b) Intensifican la toma de datos pero no modifican el plan
 - c) Son flexibles para modificar el plan y prueban ideas nuevas
11. Los científicos más destacados ¿siguen la secuencia del método científico?
 - a) No, aunque siguen un plan, son creativos y flexibles ante la sorpresa
 - b) Sí, el método asegura resultados más válidos y racionales
 - c) Sí, el método asegura la toma de datos fiables
12. Un músico es aficionado a la observación de las estrellas usando un telescopio. Tras una decena de años, ha forjado una teoría sobre la expansión del Universo. Sobre todo ¿qué debe mostrar en su publicación para que los expertos le den alguna credibilidad?
 - a) Que ha constatado su teoría muchas veces con observaciones
 - b) Que parte conociendo las demás teorías sobre el tema
 - c) Que ha observado siguiendo el rigor de un método científico
13. Un agricultor lleva una década experimentando y consigue un producto químico que combate bien cierta plaga de sus frutales. Es probable que la comunidad de expertos no admita que sus resultados se puedan generalizar porque:
 - a) No confirman sus resultados
 - b) No muestra conocer la ciencia
 - c) No ha usado el método científico
14. En general, cuando la ciencia crece por las aportaciones individuales ¿a qué se parece más?
 - a) Al encaje de una nueva pieza en un puzle
 - b) Al aumento del agua de un vaso donde se echa una gota
 - c) Al crecimiento de un ser vivo tras digerir sucesivas comidas
15. Respecto a la participación de los científicos, se puede decir que la mayoría de teorías se debe:
 - a) Al genio intelectual de pocos
 - b) A que se sigue un plan concebido por una mayoría
 - c) Al esfuerzo coordinado de muchos
16. Un científico novel quiere ser reconocido por sus estudios de un fenómeno que cree desconocido, ¿cuál será el primer paso que debe dar para iniciar la investigación?
 - a) Detallar un plan de trabajo
 - b) Observar con detalle el fenómeno
 - c) Leer lo escrito sobre el fenómeno
17. Es mejor ver la ciencia como un conocimiento que:
 - a) Propone y contrasta teorías para explicar la realidad material
 - b) Contiene un conjunto de ideas coherente y consensuado
 - c) Intenta descubrir el orden que existe en la naturaleza
18. Hacer ciencia es, sobre todo:
 - a) Buscar las leyes naturales
 - b) Confrontar teorías y medio físico
 - c) Usar con rigor el método científico
19. Sobre todo, la ciencia se distingue de otros conocimientos porque es el más:
 - a) Preciso
 - b) Contrastado
 - c) Útil
20. La razón más convincente para afirmar que no es científico lo escrito sobre ovnis es que:
 - a) No es posible probar que es falso
 - b) Es un producto de la imaginación
 - c) Nunca se han visto de cerca
21. La razón principal que explica la complejidad del conocimiento de ciencias es:
 - a) Las ideas cambian con frecuencia ante nuevos datos y debates
 - b) La acumulación de datos impide crear cierto orden
 - c) La realidad del experto cambia con el cambio de teorías

22. La ciencia se distingue de otros conocimientos porque es el más:
- Útil y eficaz al sector de bienes materiales
 - Riguroso en el uso del método científico
 - Fiel reflejando las leyes naturales
23. La ciencia ¿puede explicar cualquier problema de la realidad?
- No, sólo el de un sector ligado a lo material
 - Sólo del sector racional pero no del irracional (p.e. sentimientos)
 - Sí, pues es el conocimiento más preciso de la realidad
24. Un agricultor afronta el problema anual de tener una buena cosecha de patatas. ¿Resolvería este problema mejor un científico?
- No, las teorías de ciencias se alejan de la actividad práctica
 - Es posible, si sus conocimientos versan sobre la patata
 - Sí, su nivel para razonar es útil para todo tipo de problemas
25. El proceso que le da a la ciencia su mayor éxito es su tenacidad para:
- Buscar datos reflejo de leyes naturales
 - Comprobar empíricamente las teorías
 - Esforzarse por consensuar y racionalizar
26. La ciencia es sobre todo:
- El conocimiento más preciso y exacto que existe
 - Un montaje teórico que se ajusta a los datos empíricos
 - Un esfuerzo racional y conjunto de gente experta
27. ¿Es cierto que la ciencia progresa más desde la flexibilidad que desde la rigidez?
- Sí, la confrontación de la teoría con la experiencia no es rígida
 - No, el éxito de la ciencia se debe al rigor del método científico
 - No, datos empíricos contrastados pueden eliminar sin más una teoría
28. Dicen que «la experiencia es la madre de la ciencia». ¿Es cierto?
- No. Más bien es un esfuerzo racional coherente y consensuado
 - Sí. La ciencia se construye desde la experimentación
 - Sí. Teoría y datos siempre se están enfrentando
29. Entre teoría y datos experimentales, ¿a qué da más credibilidad la ciencia?
- A la teoría, síntesis de muchos datos
 - A los datos, fiel reflejo de la naturaleza
 - Ambos tienen semejante credibilidad
30. ¿Será cierta la frase «mayor o menor, siempre se comete error al medir»?:
- Sí, no hay manera de medir exacto
 - Es falsa a nivel microscópico
 - No, si se usan aparatos de precisión
31. Los errores que cometen los científicos en su trabajo ¿pueden retrasar el avance de la ciencia?
- Sí, en los casos en que conducen a conclusiones falsas
 - Sí, pero el uso de un método de trabajo riguroso los disminuyen
 - Es normal que haya errores donde hay progreso científico
32. ¿Deberían intentar los científicos no cometer errores?
- Sí, retrasan el avance de la ciencia y llevan a conclusiones falsas
 - Lo normal es que, si están bien formados, apenas cometan fallos
 - Los errores no se pueden evitar, aparecen en toda investigación
33. La ciencia ¿se parece en algo a la novela?
- Sí, ambas son invenciones del hombre
 - Sí, ambas pueden estar basadas en hechos reales
 - No, la primera es fiel reflejo de la realidad y la segunda es invención
34. Es más adecuado pensar que la tarea principal de la ciencia es:
- Crear un sistema coherente de ideas de utilidad universal
 - Observar y experimentar con neutralidad la naturaleza
 - Idear modelos para interpretar y controlar la naturaleza
35. Los modelos científicos ¿reproducen la realidad?
- No, son sólo invenciones contrastadas con datos
 - Sí, si están contrastados y consensuados por los expertos
 - Sí, pues se van construyendo con datos empíricos
36. ¿Es necesario que un científico tenga que convencer a sus colegas de la validez de su descubrimiento?
- Sí, si quiere que los demás reconozcan su descubrimiento
 - No será necesario si los datos empíricos hablan por sí solos
 - Sí, si pretende incorporarlo al ámbito de conocimientos científicos
37. Sobre la constitución de la materia, una nueva teoría termina siendo aceptada por los expertos en detrimento de la antigua. Lo más probable es que a partir de ahora la realidad material se:
- Vea diferente y se investigue buscando otras cosas
 - Vea igual pero se investigue con más acierto
 - Vea igual pero se investigue con datos más exactos
38. De pronto, un experimento aporta datos que contradicen la actual teoría de la luz. ¿Qué es lo más probable que revisen en un primer momento los científicos?
- La misma teoría para ver en qué ha podido fallar
 - La fiabilidad de los datos de dicho experimento
 - Aspectos de la teoría pero sin tocar su núcleo principal
39. Los datos acumulados en un año muestran que una teoría está equivocada ¿qué harán los expertos?
- la van retocando si no tienen otra teoría alternativa
 - la mantienen restando importancia a los datos empíricos
 - no se espera tanto, la rechazan con las primeras anomalías
40. Recientemente se ha anunciado en los medios de comunicación el éxito de la primera clonación humana. ¿Es usual en ámbitos científicos la rápida publicación?
- Sí, es lo usual para garantizar la primicia
 - No, antes se discute en foros científicos
 - Sí, si su implicación social requiere una publicación rápida

Development, validation and evaluation of a multiple-choice questionnaire to identify and to characterize preservice teachers' science views

MARÍN, NICOLÁS¹ y BENARROCH, ALICIA²

¹ Departamento de Didáctica de las Matemáticas y de las Ciencias Experimentales. Universidad de Almería.

² Departamento de Didáctica de las Ciencias Experimentales. Universidad de Granada.

nicolas.marin@gmail.com

aliciabb@ugr.es

Abstract

In recent years, there has been an obvious effort among researchers in science education about the epistemological conceptions of future teachers. This work fits into this line of research, and shows the design, development and validation of a multiple-choice questionnaire to evaluate scientific beliefs of secondary future teachers (COMVDC).

The main feature of the design of the questionnaire is that we have tried to fit it to the content on the nature of science that we should teach to future teachers of secondary schools. This is, after doing a literature review on the epistemological conceptions of teachers, we conclude the desirability of a new questionnaire that could be done in a hour, in the context of teacher training at secondary school. So, a pre-post-test administration may reveal the degree of learning related to the nature of science.

Specifically, this study aims to:

- a) Show the contents and topics about the construction and nature of science that might be of interest for future teachers of secondary schools, and that, accordingly, we are trying to evaluate with the questionnaire.
- b) Describe the process for the development of the multiple-choice items.
- c) Show in detail the process used in the validation of the initial questionnaire after its administration to a sample of 45 high school teachers in a CAP course developed in 2003.
- d) Discuss the strengths and weaknesses of the final questionnaire, after administering to a second sample of 50 high school teachers in a CAP course developed in 2005. The main goal in 2003 was the purification and validation of items, but in 2005, the special emphasis was on the usefulness of the questionnaire for the assessment of the teaching experience. In 2005, therefore, the questionnaire is used as the pretest and post-test. Another important difference between the two years is that the criteria for analyzing the suitability of each item in 2005 were more accurate, quantitative and consensus than those adopted in 2003.
- e) Draw conclusions about the results of the questionnaire and its potential to be introduced as a tool for both evaluation and instruction in the training of future science teachers.

Specifically, to analyze the validity of the questionnaire, in 2003, sixteen indicators were calculated by which the improvement of the items of the questionnaire was possible. In 2005, twenty-two indicators were calculated, which enabled the estimation of 6 factors. The valuation of the goodness of each item (upper, middle and lower) is achieved by adding the values of each factor. This last process, in order to increase its objectivity, was done by 4 groups of prospective secondary teachers and 2 experts.

In the paper, we show the questionnaire administered in 2005. It consists of 45 multiple choice items, of which thirteen items are low quality. Therefore, it would be more correct to exclude them to possible further analysis, for example, to calculate the differences between pretest and post-test. This analysis should be undertaken only with items of upper and middle quality, which have been more valid and discriminatory.

The feasibility study is done through the correlations between the indicators in 2003 and 2005, and the correlations between the indicators of pretest and post-2005. These correlations indicate the degree of stability or reliability of the items to give the same results to samples or to different moments of learning. In this paper, we show that all the correlations between rates similar comment a few exceptions that are significant in at least the 95% of the cases.

In general, the COMVDC and the self-applied mechanics make up a system that is converging toward a growing questionnaire that is valid and reliable.

The COMVDC has very good qualities to be applied in educational contexts, such as the ability to be administered within an hour and be integrated into the teaching process as an effective teaching tool. Specifically, the results of COMVDC on each student, allow the identification of not only the progress of learning (when comparing results between pretest and post) but also, as the options of some items structured as epistemological tendencies can be done, for example, groups of students or constructivist tendency to follow trends of individual epistemological.

Keywords

Teachers' cooperation, Teacher development, Teacher learning, Content Analysis, Professional Community.