

Análisis de contenido de las pruebas de acceso a la universidad en la asignatura de Química en Andalucía

Rosario Franco Mariscal¹, José María Oliva Martínez², Almoraima Gil Montero³

¹Alumna de doctorado, rosario.francomariscal@alum.uca.es

²Departamento de Didáctica, Área de Didáctica de las Ciencias Experimentales. Universidad de Cádiz, España; josemaria.oliva@uca.es

³Departamento de Química Física. Universidad de Cádiz, España; almoraima.gil@uca.es

[Recibido en abril de 2015, aceptado en julio de 2015]

Se estudia la tipología de preguntas formuladas en las pruebas de Química de acceso a la universidad en la Comunidad Autónoma Andaluza. Se analizaron 576 preguntas según los temas del currículum implicados y de acuerdo al sistema de categorías propuesto por Smith, Nakhleh y Bretz (2010) para exámenes de Química. Los temas más frecuentes fueron los de estequiometría, formulación-nomenclatura química y equilibrio ácido-base, que conjuntamente acaparaban alrededor de dos tercios de las preguntas formuladas. Las tres categorías de preguntas empleadas fueron: de Definición, Algorítmicas y Conceptuales, predominando principalmente las segundas, si bien con un cierto avance en los últimos años de las de tipo Conceptual aun cuando los cambios no resultaran estadísticamente significativos. Se observó una clara dependencia del tipo de preguntas de la temática concreta implicada en las mismas, siendo aquellas sobre formulación las que más se prestan a la categoría de “definición”, las de estequiometría y termoquímica a las “algorítmicas”, y las de átomo, Tabla periódica y enlace químico a las de tipo conceptual. A partir de los datos proporcionados se discute la necesidad de una reorientación de estas pruebas con vistas al futuro.

Palabras clave: enseñanza de la química; pruebas externas; PAU; Pruebas de acceso a la universidad; Preguntas de examen.

Type of test questions to higher education access in chemistry subject

In this work, the test question typology to higher education access in chemistry subject has been studied. This study was performed in the autonomous community of Andalusia. According to curriculum topics and the category system proposed by Smith, Nakhleh y Bretz (2010) for chemistry exams, 576 questions were analyzed. The most frequent topics were stoichiometry, formulation and chemical nomenclature and acid-base equilibria, which accounted for about two thirds of the questions. The three categories of questions used were: definition, algorithmic and conceptual questions were used. It was observed a predominance of the algorithmic questions, although in recent years, some progresses of conceptual questions have been observed, but not statistically significant changes. A clear dependence between the type of question and the specific issue involved was observed. Therefore, questions about formulation were mostly included in the “definition” category, question about stoichiometry and thermodynamic in the “algorithmic” category and atom, periodic table and chemical bond questions were included in the conceptual category. From the data obtained, the need for a reorientation in the future of PAU tests is discussed.

Keywords: chemical Education; College entrance exam; external testing, PAU; test questions.

Introducción

Este artículo forma parte de un estudio más amplio dirigido a caracterizar las Pruebas de Acceso a la Universidad (PAU) en la asignatura de Química y a estudiar su repercusión en la enseñanza de esa materia en cursos preuniversitarios.

En este caso nos centramos en el primero de los propósitos, analizando las preguntas planteadas en las PAU de la Comunidad Autónoma de Andalucía (CAA). Para este análisis nos motiva el interés por conocer el contenido y la estructura de dichas pruebas, al objeto de proponer mejoras para el diseño correspondiente, o para el de otras pruebas externas que puedan en el futuro implantarse para evaluar al sistema educativo o el desempeño del

alumnado al finalizar el bachillerato. Se trata de delimitar qué temas del currículum son los que tienen mayor peso en las pruebas, cuáles son los tipos de preguntas planteadas, y hasta qué punto se aprecia una evolución a lo largo de los doce últimos años.

Ambas pretensiones, analizar el tipo de preguntas y comprobar su efecto sobre la enseñanza, encuentran su intersección en la posibilidad de que una mejora de estas pruebas impulse también un cambio positivo en las prácticas de enseñanza del profesorado en ejercicio. De hecho la evaluación no es algo independiente de los procesos de enseñanza, encontrándose profundamente ligada a la toma de decisiones educativas (Linn, 1987).

La PAU puede considerarse dentro del grupo de lo que se han venido a llamar pruebas externas, las cuales, pese a la importante función que ejercen (Pérez Juste, 2007; Pedrinaci, 2003; Sanmartí, 2003), han sido entendidas habitualmente como instrumento de control y selección institucional, más que como elemento de regulación. No obstante, al lado de esta intención abiertamente reconocida, de forma indirecta ha ejercido otro tipo de influencia, regulando no solo al alumnado, a través de la motivación extrínseca que genera su preocupación por su rendimiento en ella, sino también las decisiones adoptadas por sus profesores en las aulas. De este modo, es un hecho incontestable, que la Selectividad se ha convertido en uno de los factores que más influye en la actividad del profesor (Banet, 2007), sesgando aquello a lo que prestan o no atención en sus clases, en función de lo que se suele preguntar en ella, y condicionando su forma de enseñar y evaluar en bachillerato (Oliva y Acevedo, 2005).

Resulta esencial, por tanto, realizar una cuidadosa reconsideración de aquello que debe ser evaluado al objeto de evitar que la evaluación se limite a dimensiones parciales del aprendizaje o a los aspectos que son más fáciles de valorar (Gil y Vilches, 2006). Sobre todo teniendo en cuenta que las evaluaciones externas, si están bien planteadas, pueden reorientar la enseñanza de las ciencias hacia propuestas innovadoras coherentes con las aportaciones de la investigación en didáctica de las ciencias (Tamir, 1998; Acevedo, 2005; Hernández, 2006).

Fundamentos y antecedentes

Analizar la naturaleza de una prueba de evaluación implica caracterizar la tipología de preguntas que compone la misma, aportar una imagen de su estructura global y describir los procedimientos utilizados para calificar cada pregunta y el resultado del conjunto. Refiriéndonos particularmente al primero de esos frentes, el análisis podría hacerse desde distintas perspectivas, entre ellas atendiendo a los temas específicos del currículo objeto de las preguntas o caracterizando las dimensiones del conocimiento que se evalúan en las mismas, lo que implica demarcar los tipos de aprendizajes evaluados y las demandas cognitivas exigidas.

Para encuadrar teóricamente la tipología de preguntas empleada en el artículo para analizar las PAU, hemos de remontarnos a la taxonomía de Bloom *et al.* (1956), en la que se establecía una jerarquía de niveles de menor a mayor complejidad para clasificar las distintas habilidades y destrezas relacionadas con la dimensión cognitiva del aprendizaje; a saber: conocimiento, comprensión, aplicación, análisis, síntesis y evaluación. En cierta forma, la taxonomía de Bloom puede considerarse un precedente de taxonomías que se emplean actualmente para clasificar preguntas de evaluación en química, por lo que tal vez convenga recordarla al objeto de encontrar sentido a los enfoques posteriores surgidos en este terreno. La tabla 1 presenta una síntesis de las características de cada uno de los niveles cognitivos propuestos por Bloom, particularizadas para el caso del aprendizaje de contenidos de química. En ella aparece toda una gama de complejidad en las tareas demandadas, siendo el nivel de “conocimiento”, el que plantea una menor exigencia al relacionarse con tareas de mera retención y recuerdo de

información, y las tareas de análisis, síntesis y evaluación, de menor a mayor por este orden, las que más complejidad encerrarían.

Tabla 1. Taxonomía de Bloom con ejemplificaciones de química (Bloom *et al.*, 1956).

Niveles	Características	Ejemplos
Conocimiento	Se refiere al recuerdo de información previamente aprendida en forma de datos, términos, afirmaciones, ecuaciones o definiciones, y que representan hechos, conceptos, leyes, teorías, convenios, sistemas de clasificación, etc.	Retener el valor del Número de Avogadro. Recordar la ecuación de los gases perfectos. Definición de isómero. Memorizar la Tabla Periódica.
Comprensión	Implica poder dar un sentido al conocimiento adquirido, lo que implica establecer relaciones a partir del mismo y poder transferirlo y generalizarlo.	Relacionar los coeficientes estequiométricos con número de moles y de moléculas. Interpretar las leyes de los gases en términos del modelo cinético-molecular. Establecer una conexión razonada entre configuración electrónica y propiedades periódicas.
Aplicación	Implica la posibilidad de usar, de manera deductiva, el conocimiento adquirido en nuevas situaciones a través de procesos que van más allá de la mera proyección directa del mismo.	Ejercicios de cálculo y de cambios de unidades mediante factores de conversión. Resolución de problemas numéricos. Usar la teoría de Lewis para explicar la formación de moléculas en un compuesto covalente.
Análisis	Supone descomponer un problema dado en sus partes, delimitando las relaciones latentes que existen.	Explicar variaciones en propiedades periódicas en función de la configuración electrónica. Dadas una serie de sustancias, ordenarlas por orden de mayor acidez a mayor basicidad. Dada una sustancia, señalar de forma razonada el carácter o no polar de la misma.
Síntesis	Se trata de componer o recomponer un conjunto a partir de las partes que lo integran. Para ello se ha de conectarlos, ordenarlos y organizarlos para formar un todo.	Dada una representación gráfica, sobre la variación de concentración de sustancias con el tiempo, determinar la velocidad de reacción. Resolución de problemas abiertos, más allá de la aplicación de algoritmos aprendidos. Diseñar experimentos que permitan recopilar datos sobre un tema dado. Por ejemplo, cómo saber si una sustancia X es un ácido o una base.
Evaluación	Se refiere a la capacidad para evaluar mediante procesos de análisis y síntesis. Requiere formular juicios de valor de acuerdo a criterios propios o sugeridos.	Dado una serie de datos experimentales, contrastar a partir de ellos el cumplimiento o no de una determinada hipótesis. Analizar virtudes y limitaciones de la Tabla Periódica. Seleccionar el mejor antiácido entre varios propuestos, a partir de una pequeña investigación.

Algunos estudios han evaluado la presencia de cada uno de estos niveles en las pruebas de evaluación en química. Por ejemplo, Sanabria-Ríos y Bretz (2010) evaluaron las características de las pruebas de química general de una facultad de Puerto Rico al objeto de sopesar el grado de coherencia de las mismas con los criterios con los que pretendieron elaborarse. Para ello, se realizó, entre otras cosas, una valoración ponderal de la presencia de cada uno de esos niveles de la taxonomía de Bloom en las preguntas planteadas, globalmente y para cada uno de los temas del currículum. Como resultado del estudio se constató que el nivel más frecuente, con diferencia, fue el de aplicación, seguido a gran distancia del de análisis, y posteriormente el de

comprensión y el de conocimiento, por este orden. Sin embargo, los niveles de síntesis y evaluación resultaron minoritarios. Esta tendencia se mantenía bastante estable a lo largo de los diferentes temas del currículum, con algunas excepciones. Por ejemplo, en temas como el de cinética, fuerzas intermoleculares y, muy especialmente, el de electroquímica, se potenciaban el nivel de “evaluación”, que puede entenderse como el más complejo de la jerarquía de Bloom.

Este esquema de niveles de exigencia cognitiva ha sido revisado y reformulado por diversos autores e incluso combinado con otros criterios para establecer otras formas de categorización. Por ejemplo Anderson y Krathwohl (Anderson y Krathwohl, 2001; Krathwohl, 2002) elaboraron una variante de dicha taxonomía en la que figuraban los siguientes niveles: recuerdo, entendimiento, aplicación, análisis, evaluación y creación. Como diferencias más notorias hay que citar el cambio de denominación sufrido por el nivel de “comprensión”, que pasa a denominarse de “entendimiento”, la reformulación del nivel de “síntesis”, que pasa a llamarse de “creación”, y el reajuste en el orden de la serie, cuyo nivel más complejo pasa a ser ahora ese último, aunque dentro de una secuencia que se entiende de un modo menos rígido y más flexible. Esta taxonomía ha servido, precisamente, para el diseño de variantes destinadas a analizar las preguntas de las pruebas PISA liberadas (Gallardo *et al.*, 2010) y las que se plantean en las pruebas de diagnóstico en Andalucía (Gallardo *et al.*, 2014). En ambos casos, parece que las capacidades evaluadas se corresponden mayoritariamente con las de orden más bajo (reproducción sobre todo en secundaria, aplicación), estando poco presente capacidades de índole superior (Gallardo *et al.*, 2010, 2014).

En el caso específico del aprendizaje de la química, se han propuesto distintos esquemas de clasificación de preguntas en íntima conexión con estas taxonomías. El trabajo de Zoller *et al.* (1995) es uno de los principales referentes de este tipo de estudios, quienes identificaron cuatro tipos de preguntas diferenciadas en función de las demandas cognitivas exigidas:

- *LOCS (Lower-Order Cognitive Skills¹)*. Se refiere a preguntas cuya resolución requiere solo el recuerdo de información o la mera aplicación directa de conocimiento teórico aprendido, a contextos y situaciones familiares. Pueden incluso constituir problemas a resolver mediante procesos algorítmicos que ya se conocen de antemano.
- *HOCS (Higher-Order Cognitive Skills²)*. Implican problemas desconocidos para el estudiante cuya resolución requiere la aplicación de conocimientos pero a situaciones desconocidas, o procesos de análisis, síntesis y evaluación.
- *Algorítmicas*. Son aquellas que requieren el uso de una serie de procedimientos memorizados para su resolución.
- *Conceptual*. Son aquellas que, formulándose de manera textual o a través de esquemas, requieren que los estudiantes utilicen conceptos subyacentes a las teorías básicas de la ciencia.

En relación a esta taxonomía, el propio Zoller en un estudio posterior, criticaba el alto peso que tienen aún en la enseñanza de las ciencias, en general, y de la química, en particular, las destrezas algorítmicas de bajo nivel, y apuntaba la necesidad de un salto hacia una mayor consideración de destrezas de alto nivel. Dicho salto es considerado esencial para promover un tipo de pensamiento lógico, racional y reflexivo focalizado hacia la toma de decisiones en la resolución de problemas complejos propios del mundo actual (Zoller, 2001).

¹Destrezas cognitivas de bajo nivel.

²Destrezas cognitivas de alto nivel.

A partir de este trabajo pionero de Zoller, se han desarrollado otras taxonomías, entre las que podemos destacar la de Stamovlasis *et al.* (2005) o la de Smith *et al.* (2010). En la primera de ellas se prefiere hablar de preguntas de recuerdo, algorítmicas y conceptuales, distinguiendo dos tipos, en el caso de las segundas, uno de mayor exigencia y otra más simple. Por su parte, Smith *et al.* (2010) propusieron otro esquema de clasificación de preguntas de exámenes al objeto de ayudar a los profesores a encontrar un equilibrio entre distintas facetas que se evalúan. Dicha taxonomía consiste en distinguir entre preguntas de Definición, Algorítmicas y Conceptuales. Según los autores citados, dicho esquema permite caracterizar cualquier cuestión de química general en un nivel razonablemente detallado, dado que contempla tanto el tipo de datos que se presentan al estudiante en la pregunta (texto, números, figuras), como el tipo de pensamiento que ha de emplear el estudiante para responder a cada una. Su esquema ha sido aplicado en estudios de investigación recientes, mostrando niveles de validez adecuados (ver tabla 2).

Entrando a comentar los resultados obtenidos por los referidos autores (Smith *et al.*, 2010), hemos de indicar su interés por caracterizar mediante dicho esquema pruebas de química general de la *American Chemical Society*. Como resultado del estudio constataron una mayoría de preguntas Conceptuales (alrededor del 50%), frente a una proporción menor de preguntas de Definición (sobre el 30%) y otra todavía más baja de preguntas Algorítmicas (en torno al 20%).

Tabla 2. Criterios de clasificación de preguntas de exámenes (Smith et al, 2010).

Categorías	Subcategorías
Definición (D)	<i>Recordar, comprender o aplicar una definición.-</i> A través de preguntas de respuesta abierta el alumnado ha de recordar, comprender o aplicar directamente una definición. Por ejemplo se solicita al alumnado que defina qué entienden por reacción exotérmica.
	<i>Reconocer una definición.-</i> En este caso la pregunta se formula en un formato de opción múltiple, de modo que el alumno ha de reconocer cuál es la respuesta correcta entre varias que se proponen como alternativas.
Algorítmica (A)	<i>Conversiones Macroscópico-Microscópicas.-</i> Se tratan de ejercicios que implican pasar cantidades del campo macroscópico (gramos, volumen, presión...) al microscópico (moles, átomos, moléculas...) o viceversa.
	<i>Análisis Dimensional-Macroscópico.-</i> Se tratan de cálculos de medidas y unidades con magnitudes macroscópicas como por ejemplo emplear la ecuación de gases ideales $PV = nRT$ o la fórmula del cálculo de la densidad.
	<i>Conversiones Microscópico-Simbólicas.-</i> Se tratan de conversiones estequiométricas a partir de fórmulas y/o ecuaciones químicas como por ejemplo en el cálculo del peso molecular.
	<i>Multietapa.-</i> Ejercicios y problemas que se han de resolver a través de varias etapas.
Conceptual (C)	<i>Explicación de ideas subyacentes.-</i> Preguntas sobre datos y fenómenos químicos que han de ser explicados.
	<i>Manejo de representaciones pictóricas de símbolos y ecuaciones.-</i> Se trata de traducir el lenguaje simbólico a diagramas de puntos o círculos para representar átomos y moléculas, o viceversa.
	<i>Análisis/Interpretación de datos.-</i> Se aporta información en forma de tablas, gráficos o datos cualitativos que han de ser interpretados.
	<i>Predicción de resultados.-</i> Se presenta una situación sobre la que se demanda hacer una predicción o previsión de resultados.

Por su parte, Sanabria-Ríos y Bretz obtuvieron resultados ligeramente diferentes en el estudio ya citado (Sanabria-Ríos y Bretz, 2010). Concretamente, tanto para las preguntas Algorítmicas

como Conceptuales pasaron a tener una presencia del 40%, quedando el 20% restante para las preguntas de Definición.

Puede verse que existe un consenso amplio a la hora de considerar que tanto el aprendizaje como la evaluación en química involucran componentes de naturaleza conceptual y algorítmica, además de recuerdo y destrezas de razonamiento de distinto orden. No obstante, cada autor ha hecho su propia propuesta de clasificación, lo que hace que, en muchos casos, las categorías postuladas solapen entre sí, haciendo muy complicada la tarea de comparar unas clasificaciones con otras (tabla 3).

Tabla 3. Taxonomías de preguntas de varios autores.

Zoller <i>et al.</i> (1995)	Stamovlasis <i>et al.</i> (2005)	Smith <i>et al.</i> (2010)
Algorítmico	Recuerdo de conocimiento	Definición
Conceptual	Algorítmica simple	Algorítmico
Destrezas cognitivas de bajo nivel (LOCS)	Algorítmica exigente	Conceptual
Destrezas cognitivas de alto nivel (HOCS)	Conceptual	

En este sentido, la taxonomía planteada por Smith *et al.* (2010) mantiene, como valor añadido, el intento de integrar distintas perspectivas. Así, serían las preguntas de Definición las que implicarían un mayor grado de recuerdo o retención de información, mientras que las Conceptuales serían las que incorporarían un mayor grado de demanda tipo *HOCS*, en términos de Zoller (Zoller *et al.*, 1995), y de destrezas de análisis, evaluación y creación, en términos de Anderson y Krathwohl (2001). Es precisamente este intento integrador el que, como veremos, nos ha llevado a utilizar dicho esquema como uno de los criterios de clasificación empleados en el estudio.

Se constata, como se puede apreciar, que tanto el tema de las evaluaciones externas como del análisis de preguntas de evaluación de exámenes, ocupan un papel importante en la agenda de los investigadores, existiendo abundante bibliografía al respecto. Sin embargo, dentro de los límites que nosotros conozcamos, no se han realizado estudios sobre las preguntas planteadas en las pruebas de selectividad en la asignatura de Química, de ahí nuestro interés por iniciar una línea de trabajo en esta dirección.

En el estudio que se presenta en este artículo se lleva a cabo, concretamente, un análisis del tipo de preguntas planteadas en las pruebas de acceso a la universidad en la asignatura de Química en la Comunidad Autónoma de Andalucía, valorando la mayor o menor presencia de los distintos temas curriculares, el tipo de respuesta exigida según la taxonomía de Smith *et al.* (2010) y la posible existencia de algún tipo de evolución al respecto a lo largo del tiempo.

Diseño de investigación

El estudio realizado es de tipo descriptivo y combina procedimientos cualitativos y cuantitativos de análisis de datos. Consiste en un análisis de contenido de las pruebas, caracterizando cada pregunta según sistemas de categorías elaborados previamente. Con posterioridad, a raíz del proceso de categorización seguido, se procedió a estudiar cuantitativamente, mediante análisis de frecuencias, la casuística de cada categoría encontrada para cada dimensión, así como la posibilidad de anclajes entre algunas de ellas, marcando sesgos o estereotipos de preguntas. Dicha cuantificación permitió además realizar un estudio evolutivo de la naturaleza de estas pruebas, comprobando si se han ido incorporando cambios con el tiempo.

Las cuestiones que orientaban el estudio fueron cuatro:

- 1º) ¿Cuáles son los temas del currículum que tienen mayor presencia en la PAU? ¿Se detectan especiales carencias o ausencias?
- 2º) ¿Qué tipo de preguntas son las que se plantean y en qué medida éstas resultan apropiadas?
- 3º) ¿Se aprecian relaciones de anclaje entre temas curriculares y tipos de preguntas?
- 4º) ¿Se observa alguna evolución con el tiempo en el panorama encontrado?

Para responder a estos interrogantes se han analizado todas y cada una de la preguntas planteadas en las doce ediciones de Selectividad que van desde 2002 hasta 2013, ambos incluidos, tanto de las convocatorias de junio como de septiembre, con sus opciones A y B. Ello supone un total de 288 preguntas. También se ha incluido en el análisis un número idéntico de preguntas de las pruebas de reserva, elegidas de forma aleatoria mediante un procedimiento estratificado y ponderal. La muestra resultante fue de 576 preguntas.

A lo largo del período de tiempo considerado, las pruebas de Selectividad en la asignatura de Química han mantenido una estructura invariable, con seis preguntas de respuesta abierta, cuatro de ellas con una orientación esencialmente cualitativa (cuestiones) y dos planteadas en formato de problemas cuantitativos.

Se trata de caracterizar cada pregunta según dos dimensiones; de un lado la temática curricular implicada en cada pregunta y, de otra, el tipo de respuesta demandada de acuerdo a la taxonomía de Smith *et al.* (2010) (ver tablas 2, 4 y 5). Estas dos últimas han sido escogidas como criterio de clasificación al haber sido desarrolladas como intento de síntesis de otras taxonomías diversas y por haber mostrado su validez en trabajos precedentes sobre análisis de pruebas de química general (Smith *et al.*, 2010, Sanabria-Ríos y Bretz, 2010). Además, dado el tipo de prueba que analizamos, es probable que dicha categorización sea la que aporte más información, sobre todo teniendo en cuenta la casi total ausencia de preguntas relacionadas con prácticas de laboratorio, las cuales sí requerirían otras categorías alternativas de análisis.

Tabla 4. Sistema de categorías utilizada para clasificar los temas del currículum correspondientes.

Contenido curricular
Estequiometría (T0)
Aproximación al trabajo científico, ciencia, tecnología y sociedad (T1)
Estructura atómica y clasificación periódica de los elementos (T2)
Enlace químico y propiedades de las sustancias (T3)
Energía de las reacciones químicas, espontaneidad (T4)
Cinética Química (T5)
Equilibrio químico (T6)
Ácidos y bases (T7)
Introducción a la electroquímica (T8)
Estudio de algunas funciones orgánicas (T9)
Formulación inorgánica y orgánica (T10)

Las unidades de análisis consideradas fueron las preguntas, a las que se asignaron tantos códigos como fueron necesarios al objeto de caracterizarlas de la forma más precisa posible según ambas dimensiones. En unos casos la multiplicidad de códigos resultantes fue debida a la necesaria articulación de varios contenidos o subtareas para responder a la pregunta en cuestión. En otros simplemente se debió a la propia estructura formulada para la pregunta, que aparecía subdividida en distintos apartados.

Tabla 5. Sistema de categorías utilizada para clasificar los tipos de preguntas.

Tipo de respuesta demandada (Smith et al, 2010)
Definición (D)
Recordar, comprender o aplicar una definición (D1)
Reconocer una definición (D2)
Algorítmica (A)
Conversiones Macroscópico-Microscópicas (A1)
Análisis dimensional-Macroscópico (A2)
Conversiones Microscópico-Simbólicas (A3)
Multietapa (A4)
Conceptual (C)
Explicación de ideas subyacentes (C1)
Manejo de representaciones pictóricas de símbolos y ecuaciones (C2)
Análisis/ Interpretación de datos (C3)
Predicción de resultados (C4)

La tabla 6 presenta algunos ejemplos que ilustran, a través de los códigos asignados a preguntas concretas, el resultado del proceso de categorización seguido.

Un aspecto esencial en cualquier proceso de análisis de contenidos es determinar la fiabilidad del proceso de categorización efectuado. En este sentido hay que decir que el análisis correspondiente ha sido llevado a cabo independientemente por dos investigadores, quienes consensuaron previamente los sistemas de categorías a emplear y los criterios de análisis correspondientes. El protocolo realizado siguió los cánones que se aportan en la literatura (Skjong y Wentworht, 2000).

Tabla 6. Ejemplos de preguntas analizadas con indicación de códigos asignados.

Enunciados	Códigos
<i>Ejemplo: 04JB4</i>	T9
“Defina los siguientes conceptos y ponga un ejemplo de cada uno de ellos: a) Isomería de función. b) Isomería de posición. c) Isomería óptica”.	D1
<i>Ejemplo: 04JA4</i>	T0
“Una bombona de butano (C_4H_{10}) contiene 12 kg de este gas. Para esta cantidad calcule: a) El número de moles de butano. b) El número de átomos de carbono y de hidrógeno. Masas atómicas: C = 12; H = 1”.	A1 A3 A4
<i>Ejemplo: 07JB2</i>	T3
“Para las moléculas CCl_4 , NH_3 y $BeCl_2$: a) Determine su geometría mediante la teoría de Repulsión de Pares de Electrones de la Capa de Valencia. b) ¿Qué tipo de hibridación presenta el átomo central? c) Razone si esas moléculas son polares”.	C1 C2 C4
<i>Ejemplo: 08SA5</i>	T0 T8
“Dada la reacción: $KMnO_4 + FeSO_4 + H_2SO_4 \rightarrow MnSO_4 + K_2SO_4 + Fe_2(SO_4)_3 + H_2O$ a) Ajuste por el método del ion-electrón esta reacción, en su forma iónica y molecular. b) ¿Qué volumen de disolución 0'02 M de permanganato de potasio se necesita para oxidar 30 mL de disolución de sulfato de hierro (II) 0'05 M, en presencia de ácido sulfúrico?”	A1 A3 A4

Al objeto de mostrar la fiabilidad del proceso de clasificación, las tablas 7 y 8 presentan los datos tanto del coeficiente kappa como del porcentaje de acuerdo, que son los estadísticos que suelen emplearse para cuantificar el grado de acuerdo entre jueces. Puede apreciarse en las mencionadas tablas un alto consenso entre jueces en la mayoría de los casos. Tan solo la categoría C4 (Realización de predicciones de resultados) el acuerdo fue solo moderado ($k=0,59$; 79%), lo que muestra el mayor grado de subjetividad de dicha categoría, al menos en la forma en que fue considerada. A pesar de ello, incluso en este caso puede considerarse “moderado” el acuerdo alcanzado (Landis y Koch, 1977); de hecho, el porcentaje de discrepancias solo afectaba a 1 de cada 5 casos (21%), que fueron resueltos por consenso entre los dos jueces, una vez estudiados caso por caso.

Tabla 7. Datos de concordancia entre jueces para la clasificación de preguntas por “Temas” del currículum (Coeficiente *kappa* y % de acuerdo entre jueces).

	Temas										
	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
K	0,74	-	0,98	0,66	1,00	1,00	1,00	0,93	0,88	0,85	1,00
%	90		98	98	100	100	100	99	97	97	100

Tabla 8. Datos de concordancia entre jueces para la clasificación de preguntas según la tipología de Smith *et al.* (2010).

	Tipos de preguntas								
	D1	D2	A1	A2	A3	C1	C2	C3	C4
K	0,95	1,00	0,87	0,92	0,96	0,95	0,90	0,87	0,59
%	98	100	94	97	98	97	96	94	74

Resultados y discusión

Temáticas curriculares abordadas en las preguntas

Un primer análisis consiste en clasificar las preguntas de acuerdo a la temática tratada. En la figura 1 se recoge la frecuencia relativa de cada uno de los temas. Hay que indicar que la suma de frecuencias relativas supera la unidad ya que en algo más de la cuarta parte de las preguntas concurrían contenidos relacionados con dos temas distintos.

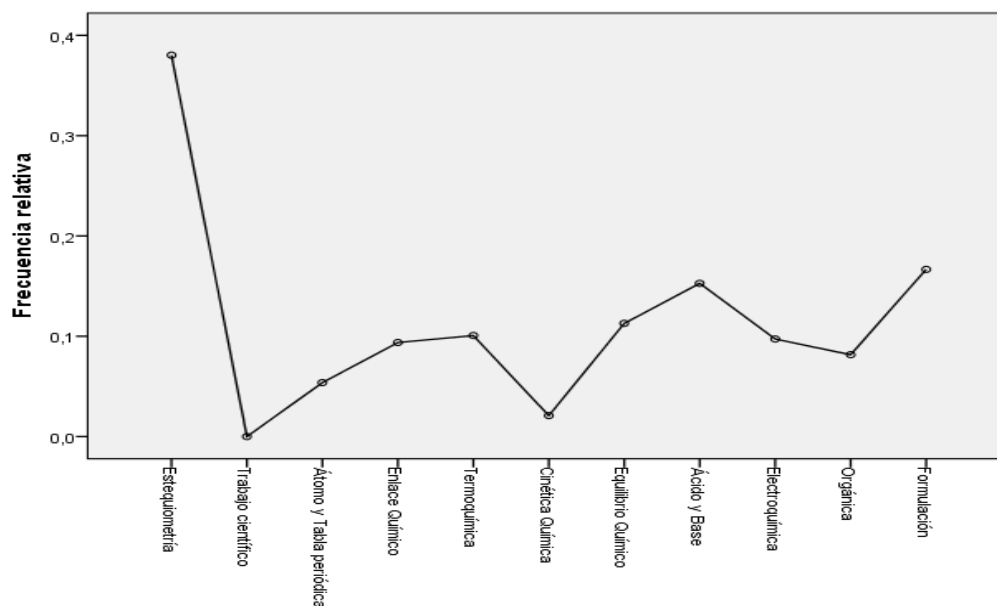


Figura 1. Frecuencia relativa de las preguntas según temática.

Se infiere de la figura 1 una temática que sobresale respecto al resto, como es la de estequiometría, en la que se acumula un total del 38% de las preguntas. Le siguen formulación con un 16,7 % y ácido y base con un 15,3 %. El resto corresponde a preguntas de otras temáticas, como equilibrio químico (11,3%), termoquímica (10,1%), electroquímica (9,7%), enlace químico (9,4%) u orgánica (8,2%). Cabe destacar que el tema del trabajo científico no aparece en ninguna pregunta, y que temas como el de cinética química o átomo y Tabla periódica aparecen de forma casi testimonial, con solo un 2,1 % y un 5,4 %, respectivamente.

Estos datos muestran un desequilibrio en la atención prestada en las pruebas a los distintos temas del currículum, y en este sentido nos parece excesiva la importancia concedida a las temáticas de formulación y estequiometría que, en conjunto, acaparan más de la mitad del peso de la prueba. Es lógico que dichos temas jueguen un papel especial, dado su importancia y su valor instrumental para los otros temas, pero dicho énfasis limita la presencia de otros temas de química que también son relevantes.

Otro aspecto a considerar es la concurrencia de temas para una misma pregunta, ya que, aunque la mayoría de ellas (74%) abarcaba solo uno, alrededor de la cuarta parte (26%) incluía dos, ya fuera por contener varios apartados o porque su resolución implicaba transversalmente contenidos de temas diferentes. Casi en todos los casos el tema concomitante con otros temas era el de estequiometría, aunque esporádicamente también aparecían combinaciones de enlace y química orgánica, ácido-base y equilibrio, o átomo y enlace químico.

Tipología de preguntas

La tabla 9 presenta los resultados de la categorización de preguntas para las categorías y subcategorías previstas por Smith *et al.* (2010).

Tabla 9. Resultados de la categorización de preguntas previstas por Smith *et al.*

Categoría	Subcategoría	Número de codificaciones	Número de preguntas**	%***
Definición	Recordar, comprender o aplicar una definición	174	150	26,0%
	Reconocer una definición	0		
Algorítmica	Conversiones Macroscópico – Microscópicas	227	245	42,6%
	Análisis Dimensional-Macroscópico	95		
	Conversiones Microscópico-Simbólicas	245		
	(Multietapa)*	(221)#		
Conceptual	Explicación de ideas subyacentes	114	181	31,4%
	Manejo de representaciones pictóricas	67		
	Análisis/Interpretación de datos	92		
	Predicción de resultados	139		
Total		1153*	576	100%
		codificaciones	preguntas	

Valores ya incluidos en las otras categorías de preguntas de tipo algorítmicas.

* Contabilizando todos los códigos asignados a cada pregunta, incluso si concurrían diferentes subcategorías dentro de una misma categoría.

** Contabilizando de forma prorrateada el número de preguntas en las que concurrían códigos de dos categorías distintas.

*** Contabilizado respecto al total de preguntas (N=576).

En la primera columna de datos se contabiliza el número de códigos totales asignados a las preguntas, con independencia de si concurrían o no diferentes subcategorías dentro de una misma categoría. En este sentido, ya que habitualmente una misma pregunta se categorizaba simultáneamente en dos subcategorías distintas, el total de codificaciones efectuadas ($N'=1.153$) fue superior al número de preguntas ($N=576$).

En la siguiente columna cada categoría se considera globalmente, recogiendo la frecuencia de preguntas asignadas a una al menos de las subcategorías correspondientes. En aquellos casos en los que concurrían varias categorías diferentes dentro de la misma pregunta, se prorrateó su aportación entre las distintas componentes. Con ello se pretendía que la suma total coincidiese justamente con el número total de preguntas, y que los porcentajes resultantes pudiesen interpretarse como el peso de la categoría en la configuración global de las pruebas.

Puede apreciarse que el mayor peso (42,6%) se corresponde con la categoría de preguntas Algorítmicas, y dentro de ellas, particularmente, la subcategoría de “Conversiones Microscópico-simbólico”. La mayoría de estas preguntas son tareas de resolución de problemas numéricos, de acuerdo a modelos-marco de resolución previamente aprendidos por los alumnos. Le sigue la categoría de preguntas Conceptuales (31,4%), y dentro de ellas las preguntas de “Predicción de resultados”. Finalmente, son las preguntas de Definición las más escasas (26,0%).

Cabe destacar, especialmente, la baja proporción de preguntas que exigen el manejo de representaciones pictóricas, a pesar de la importancia que tienen las imágenes en la modelización en Química a nivel atómico y molecular. De este modo, es posible que muchas de las dificultades que encuentran los alumnos para entender nociones y procesos en química se deban a su incapacidad para visualizar mentalmente o a través de representaciones icónicas (Bodner y Domin, 2000), aspectos como el significado de una fórmula, de una ecuación química o de un equilibrio químico. De hecho, la modelización en química se ha convertido en los últimos años en un tema importante dentro de la enseñanza de la química, y en una competencia que ineludiblemente debería desarrollar el alumnado como parte de la competencia científica (Kozma y Russell, 2005; Justi, 2009; Grilli *et al.*, 2015; Oliva, Aragón y Cuesta, 2015).

En conjunto, los datos obtenidos muestran una composición distinta a la encontrada por otros autores analizando pruebas estándares diseñadas para Química General en otros países. Así, Smith *et al.* (2010), encontraron en las pruebas diseñadas por la *American Chemical Society* una mayor proporción de preguntas de Definición, pero un porcentaje más alto también de preguntas Conceptuales. De este modo, las preguntas conceptuales, que son las más innovadoras y más exigentes en opinión de los citados autores, abarcaban el 50% aproximadamente de la prueba, frente a solo el 31% encontrado solo en las pruebas de acceso a la universidad de Química en Andalucía. Los resultados obtenidos en este estudio se aproximan más a los obtenidos por Sanabria-Ríos y Bretz (2010), quienes obtuvieron una distribución de 21%-38%-41% (definición-algorítmica-conceptual), frente al 26%-43%-31% obtenido en nuestro caso. Aun resultando similares, todavía en nuestro caso la proporción de preguntas conceptuales sigue siendo menor (figura 2).

Por tanto, los resultados del estudio muestran que, aunque las PAU de Química en Andalucía mantienen una proporción adecuada de preguntas de “Definición”, presentan un sesgo excesivo, al igual que en el caso de Sanabria-Ríos y Bretz (2010), hacia preguntas de tipo “Algorítmicas” en detrimento de las de naturaleza “Conceptual”. Estas preguntas de tipo “Algorítmicas” eran, en su totalidad, de carácter cuantitativo, y centradas en la resolución de problemas tipo. Son por tanto preguntas que podríamos clasificar como *LOCS* en los términos propuestos por Zoller *et al.* (1995). No se incluyen, sin embargo, otro tipo de

preguntas cuantitativas de carácter más abierto y orientadas a mostrar la creatividad de los estudiantes y su capacidad para resolver problemas novedosos, no necesariamente complejos.

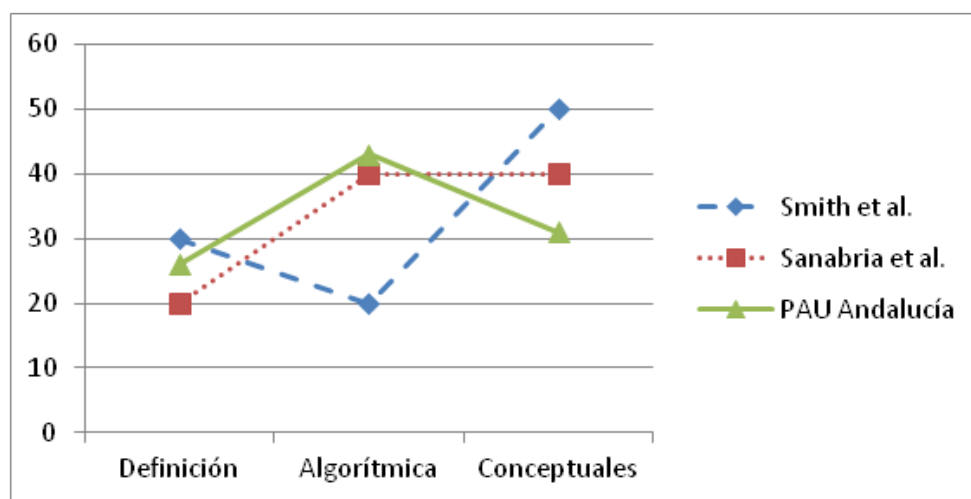


Figura 2. Comparación del tipo de preguntas en las PAU de Andalucía con el obtenido en pruebas institucionales de Química general en otros países.

En la figura 3 se muestra la distribución de porcentajes correspondiente a las tres categorías en función de la temática implicada, para lo que se ha empleado un diagrama triangular.

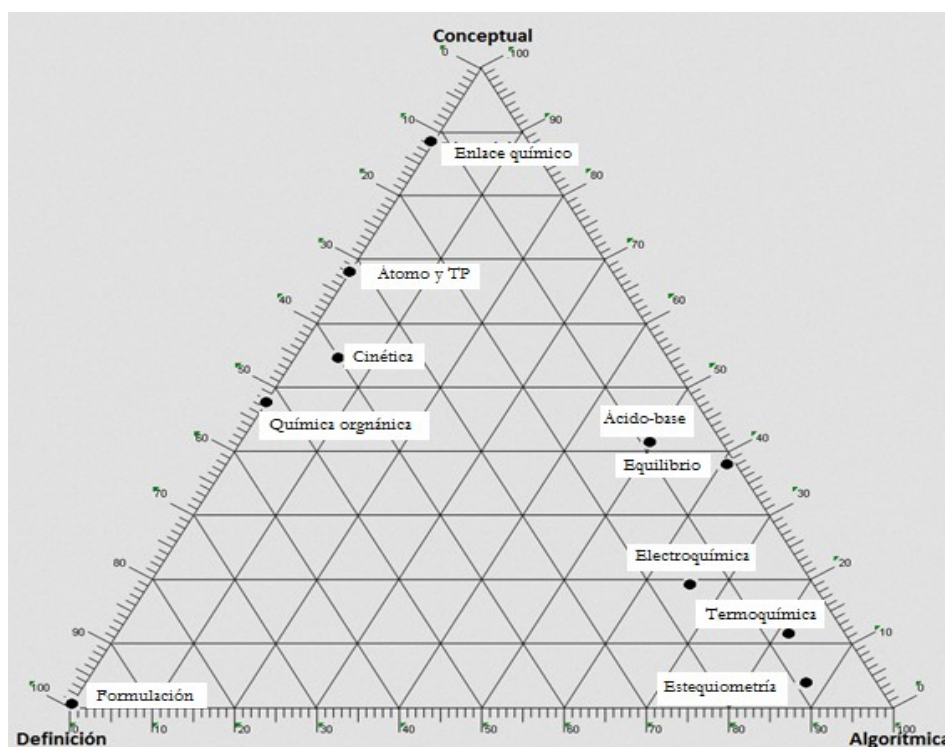


Figura 3. Diagrama trifásico para la tipología de preguntas en función de la temática curricular.

En él, la situación de un determinado tema curricular depende de tres coordenadas, que son las proporciones correspondientes a las tres categorías. Cuanto más cerca esté un tema curricular de un vértice, mayor proporción le corresponde de esa categoría y menor de las demás. Se observa una clara dependencia del tipo de preguntas, de la temática concreta

implicada en cada caso, siendo las preguntas sobre formulación las que más se prestan a la categoría de Definición, las de estequiometría y termoquímica a la de preguntas Algorítmicas, y las de átomo, Tabla periódica y enlace químico a las de tipo Conceptual.

Evolución con el tiempo de las pruebas

La figura 4 presenta una comparación entre las distribuciones de frecuencias temáticas para los tres períodos de tiempo considerados. Se aprecia una importante estabilidad a lo largo de los años, detectándose solamente cambios limitados en algunos temas. Por ejemplo, se vislumbra una pequeña reducción con el tiempo en la proporción de preguntas de estequiometría, que baja del 40% al 35%, y una ligera disminución también de las preguntas de termoquímica y orgánica. Mientras tanto, a cambio, aumentan un poco las frecuencias de preguntas correspondientes a Tabla periódica, enlace químico y cinética química. No obstante, la prueba del Chi-cuadrado no detectó diferencias estadísticamente significativas en ninguno de los casos, por lo que no podemos extraer ninguna conclusión acerca de una posible intencionalidad de los diseñadores de la prueba en relación a esa eventual evolución.

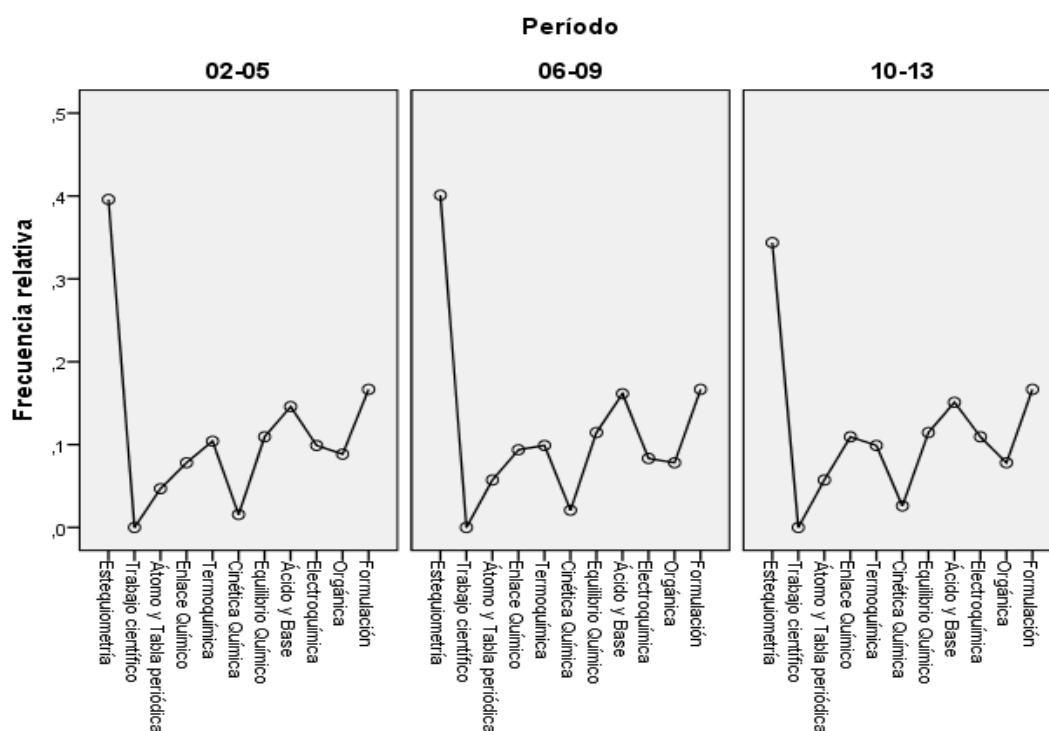


Figura 4. Frecuencia relativa de las preguntas según temática abordada por períodos.

Por su parte, la figura 5 muestra la evolución de las categorías sobre el tipo de preguntas a lo largo de tres períodos parciales en los que hemos dividido los 12 años analizados. Se aprecia con el tiempo una cierta disminución de la componente algorítmica, y levemente de la de definición, en beneficio de las de tipo conceptual. No obstante, los resultados estadísticos mediante la prueba del Chi-cuadrado tampoco en este caso mostraron una asociación estadísticamente significativa ($X^2=4,0436$; $gl=4$; $p>0,4$).

Conclusiones

Los resultados obtenidos en el estudio muestran tres sesgos importantes a la hora de configurar las PAU en Andalucía. El primer sesgo consiste en un marcado desequilibrio en el peso otorgado a los distintos temas curriculares, de modo que mientras algunos, como

estequiometría y formulación, parecen sobrevalorados, otros son objeto de una atención testimonial, como ocurre con cinética química, cuando no inexistente, como sucede con el tema dedicado al trabajo científico. Esto último, para nosotros, marca una carencia especialmente importante de las pruebas, al no tener en cuenta las tendencias actuales que valoran los aspectos relacionados con la naturaleza de la ciencia como un importante foco de aprendizaje. Tal ausencia resulta, por otra parte, coherente con la escasa atención recibida por ese tema en la enseñanza habitual en los centros escolares, algo que también merece una seria revisión.

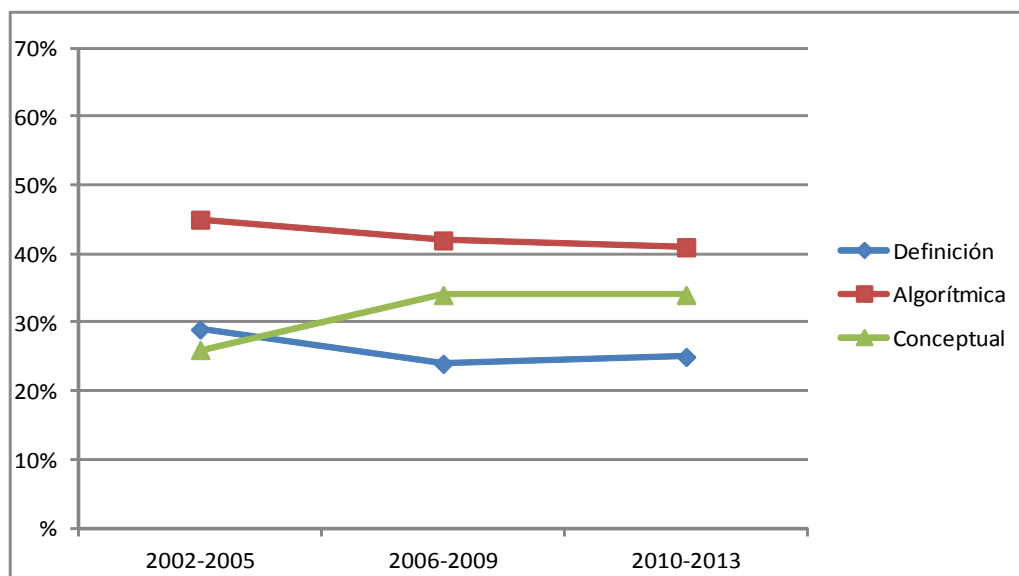


Figura 5. Evolución con el tiempo de las pruebas de acceso.

El segundo sesgo apreciado se refiere a una clara superioridad de la componente de tipo algorítmica, frente a la conceptual, de modo que mientras la primera acapara más del 40% de los códigos asignados, solo menos de la tercera parte es de naturaleza conceptual. Por tanto, puede decirse que existe una importante tendencia en las pruebas a evaluar destrezas en la resolución de ejercicios y problemas tipo, sobre todo cuando los resultados se comparan con los obtenidos en estudios como el de Smith *et al.* (2010). Esas destrezas se encuentran más relacionadas con el aprendizaje mecánico y rutinario de procesos tipificados de resolución de “ejercicios”, que con la madurez de razonamiento y la competencia científica, que quizás debían de ser el verdadero foco de una prueba de este tipo. Sobre todo se detecta una carencia de preguntas conceptuales que evalúen las destrezas del alumno para manejar representaciones pictóricas a nivel atómico-molecular, un aspecto que se ha considerado esencial para la comprensión del alumnado y para la resolución de problemas de química (Bodner y Domin, 2000). Hay que decir, en este sentido, que incluso las así llamadas “cuestiones” de la prueba, que podría esperarse que sirviesen para abundar en la componente cualitativa y comprensiva del aprendizaje, en la práctica se aprovechan a veces para plantear ejercicios de aplicación directa de cálculos químicos, confiriendo así a las mismas un carácter cuantitativo también. El dato positivo es que se aprecia en una reducción paulatina de las diferencias de proporciones entre las preguntas algorítmicas y conceptuales, si bien se detecta un cierto estancamiento de dicha aproximación en los últimos años.

El tercer sesgo que se aprecia es un efecto concomitante de anclaje de algunas de las temáticas curriculares con tipos de preguntas específicas dentro del trinomio “Definición, Algorítmica o Conceptual”. En el mejor de los casos aparecen temas que se presentan bajo dos tipos de preguntas distintas. Ello implica la existencia de estereotipos latentes tras las preguntas, con

poco espacio para otras alternativas distintas a las acostumbradas. Por ejemplo, mientras que temas como estequiometría o termoquímica se presentan casi únicamente en el formato algorítmico, otros como el de enlace químico se abordan solo de manera conceptual. Esto, que en el segundo caso citado podría considerarse totalmente lógico, dada la naturaleza esencialmente cualitativa de los aprendizajes demandados en ese tema; no lo es en el primero, en donde una parte importante del currículum exige no solo el desarrollo de destrezas de resolución de ejercicios, sino un alto grado de abstracción conceptual, así como de comprensión y manejo de los distintos lenguajes de la química.

Estos sesgos que hemos mostrado, vienen a marcar no solo limitaciones importantes en la naturaleza de las PAU, algo que sería quizás los menos importantes, sino también focos de condicionamiento para el profesor de secundaria y bachillerato, que podría tender a buscar en los estereotipos señalados, los referentes en los cuales basar sus clases. Si ello fuera así, de forma implícita, tales sesgos estarían proyectándose más allá de los límites de la propia prueba, alcanzando al quehacer diario del profesorado, ante la necesidad de preparar a los alumnos para ese tipo de pruebas.

Tal proyección es algo que suele percibirse habitualmente en los centros escolares y en las conversaciones del profesorado, habiendo sido constatado incluso en la literatura en didáctica de las ciencias (Banet, 2007). De hecho, muchos profesores declaran centrar las clases de química de 2º de bachillerato, e incluso en cursos anteriores, en la preparación del alumnado para esta prueba. De ahí que no sea de extrañar que, en la práctica, se dedique especialmente atención a la resolución de problemas tipo de carácter cuantitativo (algorítmicos), prestando una menor atención al tratamiento de aspectos conceptuales y al desarrollo de destrezas de orden superior que impliquen actividades de análisis, evaluación o de tipo creativo. También puede considerarse normal, desde esta perspectiva, la escasa atención prestada habitualmente a las prácticas de laboratorio, como tampoco a aspectos relacionados con la historia de la química, la naturaleza de la ciencia y perspectivas próximas a las orientaciones Ciencia-Tecnología-Sociedad (CTS) del currículum. Es evidente que puestos a aligerar un currículum, que ya de por sí es bastante amplio, la peor parte se la lleven aquellos temas o tipos de actividades sobre los que nada o poco se pregunta en las PAU.

No obstante, si bien es posible que las pruebas de acceso a la universidad condicionen estrechamente la actuación del profesorado, también es factible considerar un efecto recíproco, de modo que sean los hábitos y rutinas de enseñanza predominantes en las clases habituales de secundaria y bachillerato las que orienten el contenido y la estructura de los exámenes de Selectividad. Si esto fuera así, realmente tendríamos un efecto bucle, de retroalimentación, según el cual, las pruebas de acceso se confeccionarían de acuerdo a la práctica de enseñanza habitual, para estar en sintonía con lo que el profesorado espera de ella; y, a su vez, también, vendrían a marcar pautas y estereotipos que los propios profesores se sentirían obligados a seguir a la hora de implementar el currículum en sus clases de bachillerato, e incluso de cursos anteriores.

Todo ello sugiere la necesidad de una profunda reorientación en la naturaleza y estructura de estas pruebas, o al menos de las que pudieran emplearse en el futuro para regular el acceso del alumnado a la universidad o como pruebas de nivel de las distintas reválidas. Sería preciso evolucionar hacia pruebas más equilibradas que eviten sesgos como los detectados y que pongan más en valor aprendizajes de tipo cualitativo. Dicha evolución también debería implicar un cambio sustancial en la tipología de preguntas, hacia otras que, sin tener por qué resultar excesivamente complejas, impliquen otro tipo de estrategias de razonamiento de mayor nivel que la mera reproducción de algoritmos de resolución ya aprendidos, en línea, por ejemplo, con la estructura de las preguntas empleadas en las pruebas PISA. Aunque en los

últimos años se detecta una interesante evolución de las pruebas PAU de Andalucía hacia una mayor proporción de preguntas conceptuales y una menor presencia a cambio de las de tipo algorítmico, todavía dicha evolución es limitada.

Podría alguien argumentar que una mayor presencia de preguntas conceptuales que impliquen modelización a nivel submicroscópico o demandas tipo *HOCs*, corre el riesgo de conducir a pruebas más complejas que en la actualidad. De hecho, existe evidencia de que muchos alumnos que son capaces de resolver problemas de tipo algorítmicos, fallan a la hora de afrontar preguntas conceptuales, sencillamente porque no están familiarizados con tal tipo de preguntas (Zoller *et al.*, 1995; Nakheh *et al.*, 1996). Sin embargo, esos mismos trabajos parecen apuntar a que el alumnado, una vez que se habitúa en clase a afrontar preguntas de ese tipo, mejora ostensiblemente en su capacidad para resolver preguntas de esa naturaleza, aun cuando sean de mayores exigencias cognitivas. Así mismo, parecen indicar que dicha capacidad influye positivamente en las actitudes del alumnado hacia la química y en su habilidad para resolver preguntas algorítmicas (Zoller *et al.*, 1995; Nakheh *et al.*, 1996).

Pero al lado de este necesario cambio de orientación de las PAU, resulta imprescindible atacar también el problema que supone la inercia al cambio en la cultura docente y la resistencia a la introducción de innovaciones en la enseñanza; y ello solo es posible desde la formación del profesorado, especialmente en su etapa inicial en la que ya debería abordarse este problema como foco de análisis y en torno al que debatir. De hecho, suele ser un tema que el profesorado en formación suele sacar a colación para justificar la dificultad de un cambio didáctico en las aulas, precisamente adoptando como argumento la escasa sintonía que existe entre los cambios e innovaciones que se les propone desde la didáctica de las ciencias y la realidad actual de las PAU, ancladas todavía en posiciones sumamente tradicionales.

A pesar de que todos los resultados del presente estudio parecen apuntar en la dirección señalada, todavía se requiere más investigación al respecto dentro de esta línea de análisis de las pruebas PAU. En concreto, convendría avanzar en tres direcciones distintas, antes de extraer conclusiones definitivas.

De un lado, sería preciso comparar los resultados obtenidos en Andalucía con los de otras Comunidades Autónomas, al objeto de realizar un diagnóstico global en España. En este sentido, y aunque reconocemos como un límite de la investigación el haber trabajado solo con las PAU de Andalucía, podemos avanzar que un análisis somero realizado de los exámenes de otras comunidades, revelaron multitud de coincidencias con las que aquí hemos expuesto.

De otro lado, sería interesante recurrir a otras tipologías de preguntas para analizar las PAU, dado que la que aquí hemos empleado (Smith *et al.*, 2010) nos informa sobre la estructura de la respuesta demandada, y quizás también del nivel de complejidad del razonamiento exigido, pero dice poco acerca de la interacción entre el tipo de contenido curricular tratado y el nivel de demanda cognitiva pedido. En este sentido el marco de las competencias podría resultar útil para desarrollar una taxonomía más próxima a los contenidos abordados y a la estructura de la tarea implicada en cada caso. De esta forma, podríamos intentar comprobar, de una forma más directa, mediante el análisis de las competencias implicadas, el peso que aspectos tales como los conocimientos teóricos, el trabajo de laboratorio, la naturaleza de la ciencia o aspectos relativos al lenguaje de la química, entre otros muchos, poseen en estas pruebas. Si bien es verdad que en el estudio que acabamos de mostrar se presentan algunos indicadores que sugieren una pobre presencia de los mismos en las pruebas, estos resultan ciertamente de validez limitada dado el carácter indirecto que poseen. En este sentido puede ser interesante adoptar esquemas de clasificación como los propuestos por Gallardo *et al.* (2014) o por Cañas *et al.* (2014) para el análisis de pruebas de diagnóstico; el aportado por Gil-Montero *et al.* (2012) para el análisis de pruebas de nivel del alumnado de cursos de Química general, o el

ideado por Menoyo (2003) para evaluar las habilidades científicas evaluadas en las PAU, en este caso de Biología.

Finalmente, sería interesante evaluar la incidencia real de la Selectividad en la enseñanza de la química en los centros escolares, desde el punto de vista del condicionamiento que pudiera ejercer sobre el profesorado y/o el alumnado. Para ello, sería preciso consultar las percepciones al respecto, tanto de profesores como de alumnos.

A todo ello estamos dedicando parte de nuestro esfuerzo en la actualidad, y esperamos pronto poder aportar datos y conclusiones al respecto.

Agradecimientos

Los autores del trabajo desean expresar su agradecimiento a los revisores que lo han evaluado, ya que con sus críticas constructivas y sugerencias han contribuido a mejorar sustancialmente la versión inicial enviada.

Referencias bibliográficas

- Acevedo, J.A. (2005). TIMSS y PISA. Dos proyectos internacionales de evaluación del aprendizaje escolar en ciencias. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 2(3), 282-301.
- Anderson L.W. y Krathwohl D. R. (eds.) (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Banet, E. (2007). Finalidades de la educación científica en Secundaria: opinión del profesorado sobre la situación actual. *Enseñanza de las Ciencias*, 25(1), 5-20.
- Bloom B. S., Engelhart M. D., Fust E. J., Hill W. H. y Krathwohl D.R., (1956). *Taxonomy of educational objectives: part I, cognitive domain*, McKay: New York.
- Bodner, G.M. y Domin, D.S. (2000). Mental models: the role of representations in problems solving in chemistry. *University Chemistry Education*, 1(1), 24-30.
- Cañas, A.; Lupión, T. y Nieda, J. (2014). Análisis de las pruebas de evaluación de diagnóstico de ciencias de la naturaleza de 2º de la ESO en Andalucía. *Alambique*, 76, 63-70.
- Gallardo Gil, M., Mayorga Fernández, Mª.J. y Sierra Nieto, J.E. (2014). La competencia de 'conocimiento e interacción con el mundo físico y natural': Análisis de las pruebas de evaluación de diagnóstico de Andalucía. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 11(2), 160-180. Recuperado de <http://hdl.handle.net/10498/15973>
- Gallardo-Gil, M., Fernández-Navas, M., Sepúlveda-Ruiz, Mª.P., Serván, Mª.J., Yus, R. y Barquín, J. (2010). PISA y la competencia científica: Un análisis de las pruebas de PISA en el Área de Ciencias. *RELIEVE*, 16(2), 1-17. Recuperado de: http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_6.htm
- Gil, D. y Vilches, A. (2006). ¿Cómo puede contribuir el proyecto PISA a la mejora de la enseñanza de las ciencias (y de otras áreas de conocimiento)? *Revista de Educación, extraordinario*, 341, 295-311.
- Gil-Montero, A., Simonet-Morales, M., Blanco-Montilla, G. y Oliva, J.M. (2012). Analysis of Preliminary Diagnostic Tests on New Students in Scientific Degrees: a Case Study. En: G. Rodríguez-Gómez (Presidencia), *The Need for Educational Research to Champion Freedom, Education and Development for All*. Comunicación presentada en European Conference On Educational Research (ECER), Cádiz, España.

- Grilli, J.; Laxague, M. y Barboza, L. (2015). Dibujo, fotografía y Biología. Construir ciencia con y a partir de la imagen. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 12(1), 91-108. Recuperado de: <http://hdl.handle.net/10498/16926>
- Hernández Hernández, F. (2006). El informe PISA: una oportunidad para replantear el sentido del aprender en la escuela secundaria. *Revista de Educación*, número extraordinario, 357-379.
- Justi, R. (2009). Learning how to model in science classroom: Key teacher's role in supporting the development of students' modelling skills. *Educación Química*, 20(1), 32-40.
- Kozma, R. y Russell, J. (2005). Students becoming chemists: Developing representational competence. In J. Gilbert (Ed.), *Visualization in science education* (pp. 121-146). London: Kluwer.
- Krathwohl, D.R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice*, 41(4), 212-218.
- Landis J. y Koch G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-74.
- Linn, M.C. (1987). Establishing a research base for science education: challenges, trends and recommendations. *Journal of Research in Science Teaching*, 24(3), 191-216.
- Menoyo, M^aP. (2003). Evaluación de habilidades científicas en las PAU de biología. *Alambique*, 37, 58-69.
- Nakhleh, M.B.; Lowery, K.A. y Mirchell, R.C. (1996). Narrowing the gap between concepts and algorithms in freshman chemistry. *Journal of Chemical Education*, 73(8), 758-762.
- Oliva, J. M^a y Acevedo, J. A. (2005). La enseñanza de las ciencias en primaria y secundaria hoy. Algunas propuestas de futuro. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 2(2), 142-151. Recuperado de <http://reuredc.uca.es>
- Oliva, J.M^a; Aragón, M^aM. y Cuesta, J. (2015). The competence of modelling in learning chemical change: a study with secondary school students. *International Journal of Science and Mathematics Education*, 13, 751-791.
- Pedrinaci, E. (2003). Evaluación externa: un instrumento necesario. *Alambique. Didáctica de las Ciencias Experimentales*, 37, 9-18.
- Pérez Juste, R. (2007). La evaluación externa y sus implicaciones. Aspectos técnicos, prácticos y éticos. *Avances e Supervisión Educativa*, 6. Recuperado de: http://www.adide.org/revista/index.php?option=com_content&task=view&id=197&Itemid=47
- Sanmartí, N. (2003). Evaluación externa, por qué y para qué. *Alambique. Didáctica de las Ciencias Experimentales*, 37, 9-18.
- Sanabria-Ríos, D. y Bretz, S.L. (2010). Investigating the relationship between faculty cognitive expectations about learning chemistry and the construction of exam questions. *Chemistry Education Research and Practice*, 11, 212-217.
- Skjong, R. y Wentworth, B. (2000). *Expert Judgement and risk perception*. Recuperado de <http://research.dnv.com/skj/Papers/SkjWen.pdf>
- Smith, K.C., Nakhleh, M.B. y Bretz, S.L. (2010). An expanded Framework for analyzing general chemistry exams. *Chemistry Education Research and Practice*, 11, 147-153.

- Stamovlasis, D.; Tsaparlis, G.; Kamilatos, C.; Papaoikonomou, D. y Zarotiadou, E. (2005). Conceptual understanding versus algorithmic problem solving: further evidence from national chemistry examination. *Chemistry Education Research & Practice*, 6(2), 104-118.
- Tamir, P. (1998). Assessment and evaluation in science education: opportunities to learn and outcomes, En B.J. Fraser and K.G. Tobin (Eds.), *International handbook of science education*, (pp. 761-789), London, Kluwer Academic Publishers.
- Zoller, U. (2001). Alternative assessment as (critical) means of facilitating HOCS-promoting teaching and learning in chemistry education. *Chemistry Education Research and Practice in Europe*, 2(1), 9-17.
- Zoller, U.; Lubesky, A.; Nakhleh, M.B.; Tessier, B. y Dori, J. (1995), Success on algorithmic and LOCS vs. conceptual chemistry exam questions. *Journal of Chemical Education*, 72, 987-989.