



**Un análisis de evaluaciones formales en el curso masivo de
Química del CBC-UBA¹**

Bruno, Jorge; Di Risio, Cecilia

Ciclo Básico Común, Universidad de Buenos Aires (Argentina)

Fecha de recepción: 14/Jun/2017

Fecha de aceptación: 20/Sept/2017

Resumen: Como parte del proceso continuo de revisión de los procedimientos de evaluación en un curso masivo de Química, en este trabajo describimos: a) las ventajas de adoptar un sistema de calificación de ítems equivaluados con una grilla de correspondencia para exámenes parciales y b) la sistematización del análisis preliminar de resultados a fin de complementar y orientar mejoras académico-administrativas en varios niveles: grupos de trabajo puntuales, sedes específicas y generales del curso.

Palabras clave: Evaluación – Química – Curso masivo - Análisis

Abstract: **An analysis of formal evaluations in a massive chemistry course**

Considering the procedures of student evaluation in a massive course of Chemistry as an ongoing process, here we describe: a) the advantages of adopting a marking system of partial exams based on equivalent valued items and a qualification correspondence grid, and b) the use of statistical tools to provide more insight of the results, in order to improve academic-administrative tasks at various levels: specific work groups, specific

¹ Este trabajo se realizó en el marco del Proyecto UBACyT 20020150100007BA, financiado por la Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires.



teaching sites and the course as a whole.

Keywords: Evaluation – Chemistry – Massive course - Analysis

Introducción

La Cátedra Única de Química desarrolla sus actividades en 21 sedes del Ciclo Básico Común, con una distancia entre sedes extremas de más de 400 km (y se incorporan nuevas con frecuencia casi anual), cuenta con un plantel de 160 docentes auxiliares que se renueva alrededor de un 5 % cada año, dicta la asignatura durante 6 días de la semana en hasta 5 rangos horarios de 3 horas cada uno, es coordinada por 15 profesores y dirigida por un Profesor Coordinador General. Tal estructura requiere procedimientos de acción reproducibles y documentados en todas sus actividades, capaces de evolucionar en base a su propia dinámica.

Una de ellas es la evaluación formal de los estudiantes que cursan la asignatura. Con el transcurso de los años, este programa ha evolucionado (Torres y col., 2006). Y sigue haciéndolo, razón por la cual es oportuno describir las modificaciones recientes a su estructura, así como mostrar que la ampliación de los datos iniciales que proporciona son útiles no solamente para perfeccionar su calidad y la de sus instrumentos específicos (Camilloni, 1998), sino que permite identificar con precisión y ámbito de aplicación suficientes otras acciones académico administrativas de la Cátedra así como los objetivos y los objetos de investigación didáctica de los proyectos que se llevan a cabo en su seno (por ej., Ghini y col., 2012; Insinger y col., 2014).

Los exámenes parciales de Química constan de 15 ejercicios breves a resolver en un plazo máximo de dos horas reloj, en los cuales el estudiante debe volcar sus respuestas en sendos casilleros, reservando hasta dos ítems donde debe incluir los desarrollos. Hasta el primer semestre de 2011, los ítems no fueron equivalentes, habiendo entonces por examen 5 de ellos con valor de 1 punto y los 10 restantes con



valor de 0,5 punto. Los ítems de 1 punto solían requerir alguna operación analítica y/o procedimental más que los de 0,5 punto.

Surgieron aquí algunas cuestiones (concretas y otras potencialmente indeseables) que propiciaron la consideración de un cambio de modalidad para calificar los exámenes parciales de la Cátedra, como por ejemplo:

- Dadas las características de los exámenes (cada instancia se administra en tres días distintos y cada día en hasta 5 horarios distintos, con 4 variantes de examen por aula; deben confeccionarse 48 temas de 15 ítems para cada evaluación) es problemático obtener sin lugar a dudas un conjunto de ítems del doble de dificultad que los restantes, cada subgrupo de dificultad similar entre sí y lo mismo para cada subtema y cada tema. Asimismo, no hay garantía de correspondencia entre aquello que para los estudiantes es más difícil respecto de lo considerado por los autores.
- En las asignaturas del Ciclo Básico Común, cada examen parcial se aprueba obteniendo 4 puntos sobre un total de 10 posibles. Bastaría responder un mínimo de 4 ejercicios de 1 punto para aprobar un examen parcial (26,7 % del total de contenidos). Si bien este escenario no se manifestó con frecuencia, brinda la posibilidad de finalizar exitosamente el curso con muy pocos contenidos disciplinares acreditados.
- Existe un rango muy amplio de ítems acreditados para obtener una misma calificación, que resiente la motivación de los estudiantes al percibir asimetrías marcadas entre lo que respondieron correctamente y la calificación obtenida, especialmente al comparar sus producciones respecto de las de sus compañeros en la instancia de devolución.

En la sección siguiente describiremos la modalidad de calificación de exámenes parciales adoptada por la Cátedra a instancias de los autores del presente trabajo y que está vigente a la fecha.

En otro orden y luego de la administración de los exámenes, en forma oral durante los encuentros mensuales del claustro, el cuerpo de profesores informa la tasa de aprobación por sede y banda horaria y se socializan aspectos puntuales acaecidos durante la instancia. Dado que en general no se recaba otra información cuantitativa que no sea la tasa de aprobación por banda horaria, datos tales como la confiabilidad



del instrumento utilizado, el grado de dificultad de cada subtema y más aún, de cada ítem en cada subtema y su correlación con los restantes, etc. no está disponible para los profesores que deberán preparar los exámenes posteriores. Naturalmente que suele compartirse información verbal sobre algún ítem en particular, algún ejemplo que quizás no fuera conveniente repetir a futuro, algún ejercicio que sería oportuno incorporar a la guía de trabajo semestral. Y es información valiosa, expresada por docentes con amplia experiencia disciplinar y profesional. Sería prudente, en opinión de los autores del presente trabajo, complementar la valía de aquello que se informa en cierto modo inorgánicamente, con procedimientos complementarios, registrables y disponibles para la Cátedra, atendiendo a su dinámica específica. Todo lo anterior, a efectos de perfeccionar la expectativa de administración de la Cátedra Única inspirados en la norma ISO 9001 (Angelini y col., 2005).

Así, consideramos oportuno seleccionar y adoptar un mínimo útil de herramientas estadísticas tradicionales de validación y revisión de los exámenes de la Cátedra para describir y estipular un proceso documentado y reproducible a tales fines, complementando los aportes generales que efectúa el plantel docente.

Calificación de las evaluaciones parciales

En base a las dificultades observadas y potencialmente observables de un sistema de calificación con ítems de valor diferente (IVD), sugerimos que la percepción y valoración de la dificultad al momento de la evaluación se traslade desde los docentes hacia los estudiantes, únicos dueños de sus conflictos cognitivos, por vía de valorar cada ítem de manera equivalente y asignar la calificación en base a una grilla de correspondencia ítems/calificación que al mismo tiempo aumente la producción mínima indispensable para aprobar cada instancia.

Un sistema de ítems equitativos (IE) posibilita un análisis eficiente (recordando siempre la masividad del curso y la diversidad de tareas académico-administrativas del plantel docente) de la extensión de los aprendizajes, partiendo del número de respuestas correctas (y luego la identificación de las mismas, por supuesto) producidas en ambos exámenes parciales. Asimismo, el sistema IE propende a la



equidad entre producción y calificación, esto es, a igualdad (o mucha similitud) de número de ítems totales correctos, misma calificación final en la asignatura; una cuestión virtualmente imposible de lograr con un esquema IVD (véase la Tabla 2). La percepción de esta equidad es significativa en el ideario estudiantil y acompañarla es positivo para sostener su motivación. Y dado que los estudiantes suelen autorregular dicha motivación en punto a la distancia que media entre nuestros objetivos académicos y los suyos propios respecto de la aprobación de la asignatura, los valores asignados a la grilla son útiles para establecer un nexo entre ambos.

Comparamos los sistemas IVD e IE para toda combinación del número de ejercicios mínimos y máximos para obtener una calificación dada, por parcial (Tabla 1) y para la suma de ambos parciales (Tabla 2).

Tabla 1. Grilla de correspondencia del sistema IE y demandas mínimas y máximas del sistema IVD.

Calificación	Ítems correctos		
	IE	IVD	
		Mín.	Máx.
1	0, 1 y 2	0	2
2	3 y 4	2	4
3	5 y 6	3	6
4	7	4	8
5	8	5	10
6	9 y 10	7	11
7	11	9	12
8	12	11	13
9	13 y 14	13	14
10	15	15	15

Tabla 2. Rangos de ítems correctos en los sistemas IE e IVD para obtener una calificación promedio dada.



Promedio de calificaciones	Suma de ítems correctos en dos exámenes parciales	
	IE	IVD
1	0 – 4	0 – 4
2	4 – 8	3 – 8
3	8 – 13 (*)	5 – 14
4	13 – 15	8 – 16
5	14 – 18	9 – 20
6	16 – 22 (**)	12 – 22
7	21 – 24	16 – 24
8	23 – 26	20 – 26
9	27 – 28	24 – 28
10	29 – 30	28 – 30

De lo consignado en la Tabla 1, se observa que:

- En ningún caso del sistema IE los estudiantes deben producir más ítems correctos que el máximo posible requerido por el sistema IVD para obtener una calificación dada. IE requiere un número para tal propósito que se acerca o coincide con el máximo de IVD.
- En IE, los rangos de ítems correctos para otorgar una misma calificación son mínimos, mientras que en IVD podía haber hasta 6 totales de ejercicios correctos distintos para obtener una misma calificación, por influencia de los distintos puntajes acreditables.

Considerando la información de la Tabla 2,

- En el sistema IE se requieren al menos 13 ítems correctos para aprobar el curso (43,3 %) y al menos 21 para promoverlo (70 %), contra un mínimo de 8 (26,7 %) y 16 (53,4 %) del sistema IVD.
- La mayor equidad de la relación producción/calificación de IE queda puesta de manifiesto en los rangos consignados. En el peor escenario, un estudiante con 2 ítems correctos menos respecto de otro podría obtener un punto más de promedio. Las situaciones límite, es decir, que un mismo o muy similar total de ítems correctos conduzca a dos condiciones académicas distintas (indicados



en la tabla con (*) y (**)) son tan pocas en IE que resultan muy sencillas de identificar por los correctores y de analizar en conjunto por el profesor que decide en última instancia. De hecho, pueden programarse, y así lo hemos hecho, en las bases de datos de registro de calificación, donde se hacen constar tanto las calificaciones como el total de ítems correctos, de manera que sin que medie reclamo alguno de los estudiantes en tal situación, pueden identificarse de inmediato y analizarse esos casos para decidir en definitiva la condición académica correspondiente. Con el sistema IVD, la asimetría podría llegar a ser tal que un estudiante con dos ítems correctos más que su compañero promoviera la asignatura mientras que el segundo debería recurrirla. Por supuesto, siempre habrá casos límite. El sistema IE (como cualquier otro) no los puede evitar, pero los minimiza al punto de identificarlos fácilmente, en beneficio de los estudiantes.

Para evaluar si el sistema IE modificaba el porcentaje histórico de estudiantes que aprueban la asignatura en una misma sede, mismo semestre y misma banda horaria, aplicamos el test de Gauss para diferencias de proporciones: 565 exámenes IVD con 37,5 % de aprobados respecto de 486 exámenes IE con 36,9 % de aprobados. No hubo diferencias significativas a nivel de confianza $\alpha = 0,01$.

Además, realizamos el mismo test para 1482 exámenes IVD con 37,2 % de aprobados, respecto de la misma población, si se hubieran calificado bajo el sistema IE (35,1 % de aprobados). Tampoco hubo diferencias significativas, lo cual es grato en tanto las posibilidades de inequidad límite de IVD no llegaron a expresarse significativamente al momento del cambio de modalidad, al tiempo que se evidenció la no linealidad entre la evaluación de dificultad por parte de los autores, respecto de la percepción de dificultad de quienes son examinados y a favor del sistema IE.

La principal ventaja entonces reside en que las aprobaciones y promociones calificadas según IE son de mayor "calidad", en tanto tales condiciones académicas se alcanzan explícitamente dando cuenta de un mayor número de contenidos evaluados satisfactoriamente, en consonancia con el deseo de los docentes de asignaturas ulteriores en las unidades académicas de destino final de nuestros estudiantes.



Procedimientos de validación y revisión de las evaluaciones parciales

Se registraron y analizaron las calificaciones de cada ítem de *primeros exámenes parciales* administrados en una misma instancia según se indican a continuación. Los identificados entre paréntesis son representativos del total procesado y nos basaremos en ellos en lo sucesivo.

Sede 1: bandas horarias de 7 a 10 (exámenes 1A-D), de 10 a 13 (2A-D) y de 13 a 16.

Sede 2: bandas horarias de 7 a 10 y de 10 a 13 (3A-D).

El subconjunto representativo del listado anterior que utilizaremos como ejemplos corresponde a 2092 exámenes (31380 ítems). Las sedes 1 y 2 son aquellas con mayor número de estudiantes inscriptos en la asignatura. En las bandas horarias consignadas se desempeñan 35 docentes auxiliares, coordinados por los autores. En tanto docentes a cargo del dictado en aulas y correctores, también son representativos del total del personal docente auxiliar afectado a dichas tareas. El análisis de producción de los estudiantes correspondió al total de exámenes rendidos en las sedes y bandas horarias consideradas.

El tiempo insumido para la recopilación de datos y análisis posterior resultó adecuado para extenderlo a la totalidad de la matrícula semestral de la asignatura, con miras a adoptarlo como procedimiento estandarizado en cada sede y analizado globalmente por la Coordinación General. Lo anterior, a partir de planillas de cálculo de libre disponibilidad y macros específicas desarrolladas para este trabajo en la fase de registro. Las estimaciones estadísticas realizadas para determinar la confiabilidad de los instrumentos de evaluación así como las tendencias cualitativas observadas en la producción de los estudiantes se realizaron con el paquete *Psych* (Revelle, 2015) corriendo en el programa base estadístico *R* (versión 3.1.3) (R core Team, 2015), todos de libre disponibilidad.

Análisis estadístico a nivel de examen:



Por tratarse de instrumentos de evaluación calificados dicotómicamente con ítems de dificultad variable, se utilizó el estimador de confiabilidad KR-20 (Kuder y Richardson, 1937). El coeficiente KR-20 toma valores entre 0 y 1; cuando es mayor a 0,70 se considera que la evaluación es confiable.

Además del coeficiente KR-20, en la Tabla 3 se informan también, por cada tema y subtema, el número de estudiantes que los resolvieron (N), la media de calificación (sobre 15), la desviación estándar (DE) y el índice de dificultad (DFC) o *P-value*: la fracción de estudiantes que aprobó respecto del total. $0,86 \leq DFC \leq 1$: muy fácil; $0,71 \leq DFC \leq 0,85$: fácil; $0,30 \leq DFC \leq 0,70$: moderadamente difícil; $0,15 \leq DFC \leq 0,29$: difícil; $0 \leq DFC \leq 0,14$: muy difícil.

Tabla 3. Para algunos exámenes ejemplo, confiabilidad, medias, desviaciones estándar e índices de dificultad por subtema.

Examen	N	KR-20	Media	DE	DFC
1A	145	0,83	4,73	3,58	0,31
1B	145	0,89	5,01	4,23	0,35
1C	143	0,84	5,75	3,81	0,45
1D	132	0,84	5,71	3,78	0,39
2A	134	0,84	6,16	3,89	0,47
2B	132	0,80	5,15	3,37	0,32
2C	128	0,80	5,20	3,38	0,33
2D	129	0,81	5,16	3,51	0,34
3A*	72	0,88	5,37	4,14	0,33
3B*	69	0,88	6,49	4,26	0,49
3C*	68	0,86	5,93	4,03	0,47
3D*	64	0,83	5,77	3,69	0,45

*Estos datos deben tomarse con cautela ya que $N < 75$ en cada subtema.



Los datos de la Tabla 3 muestran que todos los exámenes resultaron confiables, de dificultad moderada y sin variaciones significativas de DFC (test de Gauss para $\alpha = 0,01$) entre subtemas, entre temas o entre sedes para un mismo examen global (los exámenes 1A-D, reordenados aleatoriamente, constituyeron los exámenes 3A-D).

Análisis estadístico a nivel de ítems:

En las Tablas 4 y 5 informamos, para dos de los exámenes procesados los índices de:

a) Dificultad (DFC) (aquí se trata de la proporción que resolvió correctamente el ítem, con límites aproximados de categorización ya mencionados a nivel de examen).

b) Discriminación (DSC) o *Point-biserial*: indica si quienes respondieron correctamente un ítem dado son aquellos estudiantes que obtuvieron las mejores calificaciones totales. $-1 \leq DSC \leq 1$. El valor -1 indica que respondieron bien quienes obtuvieron las calificaciones más bajas; el valor 1 indica discriminación perfecta: sólo respondieron correctamente quienes obtuvieron las calificaciones más altas. Valores superiores al rango 0,30-0,40 se consideran discriminaciones aceptables.

c) Correlación total entre ítems (ITC): provee evidencia empírica para verificar si algún ítem del examen es inconsistente con el comportamiento promedio de los ítems restantes. Valores menores a 0,20 sugieren que el ítem en cuestión debiera revisarse o en última instancia reconsiderar su inclusión en administraciones ulteriores del examen.

d) Coeficientes KR-20 corregidos al eliminar cada uno de los ítems del examen: se espera que KR-20 disminuya o no varíe respecto del promedio (Tabla 3) cuando se omite un ítem que contribuye favorablemente, o al menos no perjudica a la confiabilidad total y que aumente en caso contrario.



Tabla 4

E	Prm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1A	DFC	.18	.52	.27	.15	.39	.57	.69	.05	.29	.11	.52	.25	.25	.29	.41
	DSC	.40	.75	.60	.44	.60	.50	.50	.13	.48	.33	.65	.44	.65	.69	.69
	ITC	.50	.48	.57	.57	.37	.22	.31	.30	.59	.58	.39	.32	.67	.66	.53
	KR20	.82	.83	.82	.82	.83	.84	.84	.83	.82	.82	.83	.83	.81	.81	.82
1B	DFC	.26	.43	.34	.26	.48	.54	.71	.09	.12	.20	.48	.26	.29	.28	.36
	DSC	.69	.67	.79	.58	.73	.65	.60	.23	.18	.48	.73	.56	.81	.75	.69
	ITC	.65	.49	.68	.54	.53	.46	.40	.41	.60	.58	.54	.54	.72	.74	.61
	KR20	.88	.89	.88	.89	.89	.89	.89	.89	.88	.88	.89	.89	.88	.88	.88
1C	DFC	.49	.53	.46	.35	.54	.46	.73	.04	.37	.20	.44	.35	.16	.29	.37
	DSC	.70	.89	.68	.68	.77	.66	.38	.13	.68	.38	.60	.43	.40	.64	.51
	ITC	.42	.66	.45	.57	.57	.50	.30	.36	.53	.42	.41	.32	.46	.59	.42
	KR20	.83	.82	.83	.82	.82	.83	.84	.84	.82	.83	.83	.84	.83	.82	.83
1D	DFC	.40	.44	.56	.19	.20	.41	.75	.08	.51	.15	.44	.46	.37	.32	.44
	DSC	.53	.72	.60	.40	.47	.63	.53	.23	.58	.42	.63	.49	.81	.81	.65
	ITC	.40	.59	.39	.43	.51	.46	.40	.39	.39	.51	.41	.32	.62	.70	.50
	KR20	.83	.82	.83	.83	.83	.83	.83	.83	.83	.83	.83	.84	.82	.81	.83



Tabla 5

E	Prm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3A	DFC	.25	.47	.42	.26	.24	.33	.78	.13	.43	.24	.53	.31	.31	.31	.35
	DSC	.58	.58	.75	.67	.58	.50	.46	.21	.83	.67	.75	.58	.75	.71	.50
	ITC	.54	.35	.50	.69	.66	.49	.32	.26	.68	.75	.53	.47	.71	.64	.46
	KR20	.87	.88	.87	.86	.87	.87	.88	.88	.86	.86	.87	.87	.86	.87	.87
3B	DFC	.12	.65	.26	.33	.54	.55	.77	.20	.61	.29	.62	.54	.32	.32	.38
	DSC	.35	.65	.57	.74	.57	.78	.61	.43	.65	.78	.83	.78	.83	.74	.48
	ITC	.43	.52	.54	.61	.48	.55	.49	.44	.48	.72	.57	.52	.72	.69	.41
	KR20	.88	.88	.88	.87	.88	.88	.88	.88	.88	.87	.88	.88	.87	.87	.88
3C	DFC	.41	.47	.57	.35	.47	.56	.76	.19	.26	.10	.47	.26	.38	.34	.51
	DSC	.72	.77	.77	.68	.77	.72	.68	.23	.41	.23	.41	.45	.91	.82	.73
	ITC	.52	.54	.58	.57	.57	.45	.51	.31	.40	.37	.29	.33	.69	.70	.58
	KR20	.85	.85	.85	.85	.85	.85	.85	.86	.86	.86	.86	.86	.84	.84	.85
3D	DFC	.51	.69	.44	.22	.44	.61	.41	.19	.09	.20	.55	.36	.28	.25	.34
	DSC	.67	.57	.95	.52	.95	.31	.24	.43	.19	.38	.71	.67	.71	.62	.57
	ITC	.47	.43	.71	.49	.69	.24	.19	.37	.24	.33	.48	.48	.65	.54	.45
	KR20	.82	.82	.80	.82	.80	.83	.84	.82	.83	.83	.82	.82	.80	.81	.82

En las Tablas 4 y 5, así como para los casos no informados, observamos que todo aumento del coeficiente KR-20 corregido por omisión de un ítem ocurrió en simultáneo con un ITC menor a 0,20 en el mismo ítem. En tanto ambos son estimadores de la influencia de cada ítem con relación a la confiabilidad del instrumento en su conjunto y que ITC muestra un rango más amplio de valores, nos referiremos solamente a éste al momento de considerar cualquier revisión eventual de consignas.

Para ejemplificar unas pocas consideraciones, tendremos en cuenta algunos parámetros que hubieran quedado por debajo de los rangos mínimos adecuados para validar su presencia en el instrumento de evaluación. Por supuesto que las posibilidades de análisis que surgen a partir de la totalidad de datos tabulados son



mucho más amplias y no se limitan a la mera identificación de coeficientes fuera de los rangos deseables.

Ejemplo 1

En el ítem 8 surge que hubo baja discriminación porque muy pocos estudiantes respondieron correctamente al mismo:

"Escribir las fórmulas semidesarrolladas de los aminoácidos que originaron el dipéptido siguiente: $\text{H}_2\text{NCH}_2\text{CONHCH}(\text{CH}_2\text{SH})\text{COOH}$."

Las enunciaciones en los subtemas fueron todas similares, cambiando la molécula a considerar.

Este ítem corresponde al tema "Moléculas de interés biológico" y es el último que se desarrollaba en el curso hasta una semana antes de administrar el primer examen parcial. Los estudiantes entrevistados manifestaron disponer de poco tiempo para estudiarlo adecuadamente. Dado que no resultaba posible reestructurar el cronograma de la asignatura por limitaciones de tiempo para cubrir el Programa de Contenidos, la Cátedra Única coordinó y consensuó con las Cátedras de Biología del CBC (que ya tenían el tema aludido como parte de sus Programas) una optimización de enseñanza de estos contenidos, quedando para Química los aspectos más introductorios y elementales, con menor demanda temporal. De esta forma, no se perjudicó la calidad académica de la enseñanza ni la extensión de los contenidos a desarrollar en conjunto.

Ejemplo 2

El ítem 7 del examen 3D (igual al ítem 7 de 1A) discriminó pobremente, no correlacionó bien con el resto del examen ni contribuyó a la confiabilidad general del mismo en la sede 2. No se observa lo mismo en la sede 1.

Se trata de una consigna de opción múltiple con 3 alternativas, sin requerimiento de justificación:

"Indicar cuál de los siguientes compuestos será más soluble en agua: ..."



Para el ejemplo que nos ocupa, las opciones fueron: dimetil éter, 1-bromopropano y propeno. Los 3 subtemas restantes incluyeron: a) un éster, un ácido carboxílico y un cloroalcano; b) un alcohol, un cloroalcano y un alqueno y c) un alcohol, un éter y un alquino, todos los anteriores con número de carbonos similares.

La DFC ideal para una consigna de opción múltiple con 3 elementos es de 0,67 (un valor "promedio" entre la chance azarosa, 0,33, y el puntaje ideal, 1,00).

Observamos que en 7-3D hay una resolución correcta cercana al puro azar mientras que en 7-1A la proporción es cercana a la ideal, y lo mismo para los subtemas restantes en ambas sedes.

En esta primera asignatura de química universitaria, el análisis de solubilidad se restringe a la consideración de interacciones intermoleculares entre sustancias distintas. Analizando la constitución de las opciones, sólo en 3D (1A) aparecen dos sustancias polares que no forman enlaces de hidrógeno cuando están puras en estado líquido (que es el estado de agregación bajo el cual se ilustran estos contenidos en el curso) y una esencialmente no polar. Los resultados sugieren como mínimo que quienes respondieron a esta consigna en 7-3D no consideraron interacciones cruzadas.

Por otra parte, quienes respondieron correctamente en los subtemas restantes podrían no haberlas considerado tampoco. La elección de tríadas en esos casos habilita la puesta en juego de la concepción alternativa nada infrecuente por la cual, de las posibilidades ofrecidas, aquella que tenga "el mismo tipo de interacciones intermoleculares que el agua" será automáticamente la más soluble. La información disponible no permite avanzar más allá de una hipótesis, pero precisamente es su disponibilidad la que sí provoca su validación o rechazo mediante un análisis más profundo con herramientas complementarias.

Sin perjuicio de lo anterior, la diferencia con los resultados de 7-1A sugiere realizar la constatación de las acciones didácticas en la sede 2 y en esa banda horaria, al



momento de enseñar conceptos elementales para comparar la solubilidad en agua entre sustancias moleculares sencillas.

Asimismo, las consideraciones descritas serían útiles para los autores de estos exámenes, a efectos de seleccionar los ejemplos a considerar para que las respuestas, si no requieren justificación como en este caso, evidencien mejor el análisis que los estudiantes pudieron haber realizado al momento de responder.

Ejemplo 3

Los ítems 9-3D y 9-1B (idénticos entre sí) resultaron mucho más difíciles (y poco discriminadores) respecto de, por ejemplo, los ítems 9-3B y 9-1D (idénticos entre sí). La enunciación inicia con una información general, que se aplica a los ítems 9 y 10 (cuyas consignas no dependen del resultado del ítem anterior), cada uno de los cuales se redactaron en sus casilleros correspondientes. La respuesta no requiere explicitación de desarrollos. El dato de constante de Avogadro es un distractor.

- 9-3D / 9-1B: General: "En 15,0 g de una sustancia de fórmula molecular CH_3XH hay 1,25 moles de átomos de hidrógeno. Datos: $N_A = 6,02 \times 10^{23} \text{ mol}^{-1}$."
Ítem: "Calcular la cantidad de etano (C_2H_6) que contiene la misma cantidad de átomos de hidrógeno que 15,0 g de CH_3XH ."
- 9-3B / 9-1D. General: "La densidad del nitrobeneno ($\text{C}_6\text{H}_5\text{NO}_2$) a 20°C y 1 atm, es $1,20 \text{ g/cm}^3$. Datos: $N_A = 6,02 \times 10^{23} \text{ mol}^{-1}$."
Ítem: "Calcular el volumen que ocupan 1,60 moles de nitrobeneno."

Una primera consideración sobre la mayor dificultad de 9-3D remite a aquella muy extensamente documentada para operar, por parte de los novatos, en las escalas macroscópica y nanoscópica (por ej.: Johnstone, 1991; Galagovsky y col., 2003). El enunciado de 9-3B refiere a magnitudes macroscópicas y la única alusión nanoscópica está oculta en la definición de la unidad de cantidad de sustancia.

Apoyan esta consideración que desde el punto de vista matemático, la operatoria de 9-3D es más corta y sencilla que la de 9-3B, y sin embargo el primer ítem fue respondido



correctamente por menos de un tercio de aquellos que tuvieron éxito en el segundo y que, si bien las enunciaciones remiten a consignas bien distintas para subtemas de un mismo examen, los ítems 10 (no detallados aquí) aluden también a magnitudes atómicas y moleculares, con DFC bajos y comparables entre subtemas.

Un aspecto no menor, pero cuya extensión no es objeto del presente, es que más allá de otras consideraciones, muchos estudiantes que resolvieron mal 9-3D / 9-1D posiblemente no comprendieron lo que se les pedía. Esto fue manifestado explícitamente por algunos de ellos a quienes indagamos informalmente en oportunidad de la etapa de revisión personalizada. También lo observamos tanto en nuestros propios cursos como en los talleres de consultas que ofrecemos en horarios extra-clase. La demanda por establecer un nexo (en este caso una magnitud en común) entre dos sustancias no pudo ser superada, ni siquiera planteada por quienes trajeron estas inquietudes a la consulta.

Sin perjuicio de lo anterior, hemos de convenir que la consigna en sí misma, similar a varias que estaban presentes en la guía de ejercitación de la Cátedra no es representativa de situaciones a las que se enfrenta un profesional químico. Tiene el mérito de ser anticipatoria de otras que son parte de la asignatura y que requieren la identificación de una magnitud común a dos sistemas para operar luego con ella (ej. procesos de dilución de soluciones, sistemas gaseosos, etc.) Sin embargo, no parece prudente adelantar esta operatoria en contextos macro-nano que se imparten en etapas iniciales del desarrollo de clases. A raíz de estas consideraciones y a la fecha, se han retirado este tipo de consignas de la ejercitación de magnitudes atómico-moleculares y consecuentemente de los instrumentos que evalúan dicho tema.

Conclusiones

La revisión de métodos de calificación y de registro y análisis detallados de las producciones estudiantiles del programa de evaluación de la asignatura, son parte del desarrollo general de la Cátedra Única en su contexto distintivo de masividad. Suman al carácter formativo de la evaluación que ya se expresa a través de la instancia de devolución personalizada.



La calificación de exámenes con ítems valuados de manera equivalente y aplicación posterior de una grilla de correspondencia, explicita un mayor requerimiento mínimo de acreditación de conocimientos, con mayor equidad y mayor libertad de acción al momento de preparar los instrumentos de evaluación, sin perjuicio para los estudiantes ni para las tareas docentes asociadas a su administración.

El registro a nivel de subtemas de examen e ítems (individuales o por bloques de contenidos) por subtema y su análisis proveen una herramienta factible de implementación, sistemática, rigurosa y documentable para identificar problemáticas o complementar y confirmar aquellas identificadas por el plantel docente durante el desarrollo natural de la actividad académica. Los datos así analizados permiten ajustar preliminarmente el alcance de las acciones superadoras que se propongan, sean éstas a nivel de redacción de exámenes (ejemplos 2 y 3), de perfeccionamiento de materiales didácticos (ejemplo 3), de revisión de acciones didácticas en sedes específicas (ejemplo 2) o bien en el total de sedes (ejemplo 3), de estructura general del Programa de Contenidos (ejemplo 1) y en general, como resultado de las investigaciones didácticas que se desarrollan al interior de la Cátedra.

Referencias

Angelini M., Baumgartner E., Guerrien D., Landau L., Lastres L., Sileo M., Torres N. Y Vázquez I. (2005). "Calidad en educación: una meta posible. Adaptación de las normas ISO a la gestión de un curso universitario". *Educación Química* 16(1): 68-72.

Camilloni, A. (1998). "La calidad de los programas de evaluación y de los instrumentos que lo integran", en: *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires, Paidós.

Galagovsky L., Rodríguez M., Stamati N. Y Morales L. (2003). "Representaciones Mentales, Lenguajes y Códigos en la Enseñanza de Ciencias Naturales. Un Ejemplo para el Aprendizaje del Concepto de Reacción Química a Partir del Concepto de



Mezcla". Enseñanza de las Ciencias 21(1): 107-121.

Ghini A., Veleiro A., Bruno J., Guerrien D., Rusler V. Y Di Risio C. (2012). "Abordaje de la Química del Carbono en un curso introductorio y masivo de Química General". Educación Química 23(3): 370-374.

Insinger K., Bruno J.A.O. Y Di Risio C.D. (2014). "Talleres con relevancia social para integración de contenidos curriculares de Química-CBC". Revista Electrónica de Didáctica en Educación Superior, N° 8 <http://www.biomilenio.net/RDISUP/portada.htm>. ISSN 1853-3159.

Johnstone, A. H. (1991). "Why is Science Difficult to Learn? Things are Seldom what They Seem". J. Computer Assisted Learning 7: 75-83.

Kuder, G. F. Y Richardson, M. W. (1937). "The theory of the estimation of test reliability". Psychometrika 2(3): 151-160.

R Core Team (2015). "R: A language and environment for statistical computing". R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Revelle, W. (2015). "Psych: Procedures for Personality and Psychological Research", Northwestern University, Evanston, Illinois, EEUU, <http://CRAN.R-project.org/package=psych> Version = 1.5.1.

Torres N., Sileo M., Landau L., Angelini M., Guerrien D., Lastres L., Vázquez I. Y Baumgartner E. (2006). "Evolución de los procedimientos de evaluación en un curso masivo de Química". Educación Química 17(4): 424-434.

