

MÉTODOS DE EVALUACIÓN EN LOS EXÁMENES JURÍDICOS

LUCIO PEGORARO

Catedrático de Derecho Público Comparado

Universidad de Bolonia (Italia)

Traducido por M^a REMEDIOS ZAMORA ROSELLÓ.

Doctora en Derecho. Derecho Administrativo

Universidad de Málaga (España)

Resumen: El artículo analiza los modelos aprendizaje y de evaluación de los estudiantes en las disciplinas humanísticas y, sobre todo en las jurídico-públicas, poniendo de relieve cómo el test de elección múltiple, utilizado en muchos países, garantiza ciertos requisitos, como la objetividad, fiabilidad, validez y la selección. Posteriormente se ilustra una iniciativa editorial desarrollada por un grupo de profesores italianos que han elaborado varios volúmenes sobre derecho interno y comparado con la intención de elaborar tests diseñados para evaluar, y también para ayudar a los estudiantes en la fase de comprobación de su aprendizaje.

Abstract: The article analyses the different types of learning processes and assessment in social studies and, in particular, in the field of public law. It then underlines that multiple-choice tests, which are used in many countries, ensure that assessment methods respect a series of conditions, i.e. those of objectivity, reliability, validity, and selectivity. It then moves on to describe an editorial initiative by a number of Italian Professors, who have authored Italian and Comparative Public Law test-books with the aim of not only providing an assessment instrument, but also to help student to self-test their grounding.

Palabras clave: Docimología, tests objetivos, Derecho Público, evaluación, iniciativas editoriales.

Keywords: Study of Testing Techniques – Objective Testing – Public Law – Assessment Methods – Editorial Initiatives.

Sumario: 1.Introducción. 2. Las técnicas de evaluación. 2.1. Las pruebas tradicionales. - 2.2 Metodologías objetivas. 2.3. Planificación de un test. 2.4. Preparación y selección de los *items*. 2.5. La corrección del test. 2.6. Evaluación. 2.7. Operaciones finales. 2.8. Verificación. 3. Materiales didácticos para las disciplinas de Derecho Público: 3.1. Algunas iniciativas de editoriales italianas. 3.2. Los tipos de *item* contenidos en los volúmenes. 3.3. Disposiciones prácticas para la aplicación de las pruebas. 3.4. Experiencia práctica. 4. Conclusiones.

1. Introducción

Durante décadas, la profunda metamorfosis sufrida por la Universidad después de la era de las protestas, ha convertido en una “oficina de exámenes” lo que antes era un lugar de aprendizaje, de diálogo, de investigación y de evaluación de lo aprendido.

Paradójicamente, la actividad que prevalece hoy en día -precisamente la evaluación de lo aprendido- normalmente se desarrolla con criterios aproximados.

Sea cual sea el sistema adoptado, a excepción de una vaga percepción basada en la experiencia, a menudo no existe conciencia de las funciones propias del examen, que la docimología -la ciencia de la evaluación- resume en las siguiente: diagnóstico; juicio sobre la validez de un plan de estudios; la predicción de las posibilidades de éxito; evaluación de los resultados obtenidos¹.

2. Las técnicas de evaluación

2.1. Las pruebas tradicionales

Las universidades italianas (y también un destacado número de no italianas), han empleado hasta no hace mucho tiempo las pruebas de verificación de aprendizaje tradicionales (prueba oral y/o de tema escrito) sin ni siquiera una voz autorizada que desafiara este sistema, a pesar de su carácter empírico y aproximado que no siempre permite a los docentes, aún siendo voluntariosos y concienzudos, ser lo más imparciales posible. Estas pruebas se han empleado especialmente en las disciplinas humanísticas y sociales y, de forma muy destacada, en el ámbito que aquí nos interesa: las ciencias jurídicas.

De las muchas críticas que comúnmente se plantean a los exámenes orales y escritos se puede destacar que, en todo caso, estas pruebas carecen de objetividad.

En el caso de las pruebas orales, las preguntas varían de un candidato a otro: el tiempo es diferente y también lo es el estado de ánimo del examinador. A ellos podemos añadir las influencias positivas o negativas de la experiencia previa en el trato con el mismo examinador (o, en la Universidad, las influencias de los buenos o mediocres resultados obtenidos en otras materias del plan de estudios). Los propios estudiantes actúan en condiciones diferentes, y muchos de ellos están obligados a soportar la tensión de la espera interminable y de las referencias que reciben de otros compañeros.

Las pruebas escritas tradicionales gozan de un cierto grado de objetividad, ya que son iguales para todos y se desarrollan en idénticas condiciones ambientales; pero tampoco resultan suficientemente fiables debido a la falta de criterios precisos en la fase de evaluación.

Desde hace ya tiempo, numerosos estudios han destacado la gravedad de esta situación: por ejemplo, ha quedado ampliamente demostrado que el mismo examen escrito puede recibir diferentes evaluaciones por correctores diferentes, e incluso del mismo evaluador en momentos sucesivos. También se señaló que a veces un profesor utiliza una amplia

¹ Entre otros, D.L. NUTTAL, *Come si pianifica un esame*, en B. HUDSON (cur.), *Introduzione alle tecniche di valutazione*, trad. it., Bologna, 1975, p. 3.

gama de evaluación, mientras que otros centran sus valoraciones sobre dos o tres clasificaciones².

En defensa de las pruebas tradicionales se recuerda que todavía tienen el requisito de colegialidad. Este requisito puede, ciertamente, nivelar las valoraciones, suavizar los excesos, pero a menudo aumenta el grado de equidad y justicia distributiva.

Se argumenta que las pruebas tradicionales también distinguen a los estudiantes mejor preparados de aquéllos que están menos preparados; pero es precisamente en esa amplia zona intermedia donde se debe diferenciar, y es en este rango de normalidad donde suelen apreciarse los errores de evaluación.

No es que las pruebas tradicionales deban ser prohibidas: debería ser más viable y eficaz el cumplimiento de determinadas normas y, sobre todo, se pueden vincular con otras herramientas de evaluación.

2.2. Metodologías objetivas

Para hacer frente a esta situación poco satisfactoria (y también por razones prácticas) se han experimentado soluciones alternativas en lugar de las pruebas tradicionales de evaluación y examen. Estas metodologías, siguiendo las indicaciones de los expertos, debían tener las mismas características que se emplearon en la preparación de las pruebas de aptitud desarrolladas al principio del siglo XIX, especialmente en los países anglosajones, a saber:

a) Objetividad: las pruebas, o tests, son iguales para todos los examinados y tienen lugar en el mismo contexto temporal y bajo idénticas condiciones ambientales; se corrigen de la misma manera y son evaluados con la misma escala;

b) Fiabilidad: los ensayos son fiables debido a que sus resultados no están sujetos a variables no deseadas. Entregados en dos etapas sucesivas en la misma muestra (grupo), los resultados son iguales; dadas dos muestras paralelas (dos grupos de sujetos que pertenecen a la misma población), se obtienen medias y distribución de puntos similares;

c) Validez: los tests de aprendizaje pueden, si se preparan con habilidad y atención, proporcionar una indicación clara sobre la preparación de los candidatos;

d) Selección: las pruebas permiten el examen de los resultados a través de una amplia gama de evaluaciones. Los resultados, en otras palabras, se distribuyen de acuerdo a una curva simétrica, en "campana" (o "gaussiana"): se concentran en las zonas medias, y con menos probabilidades de aparecer según se alejan gradualmente de la media en ambas direcciones.

² Sobre esta materia, H. PIÈRON, *Examens et docimologie*, Paris, 1963, p. 15; O. ANDREANI DENTICI, *Abilità mentale e rendimento scolastico*, Firenze, 1968, p. 137 ss., con amplias referencias bibliográficas.

2.3. Planificación de un test

Una vez tomada la decisión de utilizar un test (solo o junto con una prueba tradicional, o para la admisión a la misma), tanto para una evaluación informal como para un examen formal, debe ser identificado de antemano el propósito de la prueba, que puede ser que el examinado demuestre: a) sólo los conocimientos obtenidos, b) también el dominio -o el uso apropiado- de las operaciones mentales (deducción, juicio, generalización, etc., coesenciales para un verdadero conocimiento).

También se debe establecer el grado de dificultad atribuida a la prueba, en función del objetivo a lograr (evaluar de forma selectiva y gradual, seleccionar a los mejores, o suspender a los peores).

A este respecto hay que decir que la prueba es más fiable con una combinación adecuada de preguntas (*items*): fáciles (25%), de dificultad media (50%), difíciles (25%). Sólo en este caso se obtiene una amplia y simétrica distribución de los resultados, que objetivamente discrimina los distintos niveles de rendimiento. Si todas las preguntas son fáciles, se establece una distribución oblicua, en la que los resultados se agrupan en torno a valores altos. En el caso de que todas las preguntas sean difíciles, se produce el fenómeno inverso (la mayoría de los resultados se reúnen en torno a un valor bajo). Por último, si todas las preguntas son demasiado difíciles, se puede obtener una distribución plana (o "rectangular") con la misma frecuencia para cada puntuación, prevaleciendo sobre las respuestas un verdadero riesgo (en otras palabras, las respuestas al azar ocasionan que cada *item* reciba un número más o menos igual de respuestas correctas). Sólo un pequeño número de estudiantes excepcionales puede alcanzar una calificación positiva: por lo tanto tales pruebas sólo estarían justificadas para una selección rigurosa, y ni siquiera se consideran adecuadas en este caso.

Al organizar el test también se debe determinar la extensión de la prueba, teniendo en cuenta que es más fiable mientras más tiempo ocupe, aunque también es necesario tener presente el problema de la fatiga, y del tiempo promedio que se requiere para completar las respuestas.

Por último, debemos decidir si la prueba se desarrollará en un tiempo determinado (como sucede en la mayoría de los casos) o sin límite de tiempo.

2.4. Preparación y selección de los *items*

En la preparación de los *items*, primero se debe decidir qué tipo de preguntas se desea adoptar. Se pueden presentar diferentes opciones: completar una frase, rellenar los espacios en blanco de una frase, trabajar sobre dos series de conceptos, poner frases en orden, seleccionar definiciones, etc.

Los dos modelos más utilizados son los siguientes: a) pregunta con respuesta libre; b) pregunta con respuestas de opción múltiple.

Excepto la última -pregunta con respuestas de opción múltiple- todos los modelos que acabamos de mencionar presentan problemas más o menos complejos de corrección y evaluación.

En general -básicamente en todo el mundo- se ha optado por el test compuesto por *items* de respuesta múltiple, con dos opciones (verdadero/falso) o más alternativas, una de las cuales es la respuesta correcta y la otras son erróneas en mayor o menor grado.

Acerca de las alternativas erróneas hay que señalar que no deben ser descaradamente absurdas o ilógicas (serían descartadas de inmediato), deben tener alguna apariencia de verdad o ser plausibles. Esto lleva a reflexionar y a no dejarse engañar, es un estímulo para profundizar en un razonamiento o a controlar un dato.

Volviendo al problema de la preparación (o elección) de los *items* (tipo, grado de dificultad, número), se deben añadir algunas referencias sobre su formulación: las preguntas -breves, sencillas, claras- no tienen que incluir ningún tipo de sugerencias (es decir, expresarse en un tono neutro). Al igual que las respuestas.

2.5. Corrección del test

El método más simple de corrección consiste en asignar un punto por cada respuesta correcta. Cuando no se utilicen los sistemas informáticos, se puede emplear una cuadrícula en la que esté marcada una señal en cada respuesta correcta. Se cuentan las respuestas correctas y se otorga una puntuación provisional.

La puntuación provisional se utiliza para aplicar una penalización en el número total de errores. La deducción es de un punto por cada error en el caso de elementos con dos alternativas, un punto por dos errores en el caso de tres alternativas, un punto por cada tres errores cuando las alternativas son cuatro, etc.

Las omisiones se obvian. Ya han castigado el resultado global, restando puntos potenciales (de hecho, respondiendo al azar se puede llegar a obtener algún punto, sobre todo si la incertidumbre no es total y sólo afecta a una parte de la respuesta). Al mismo tiempo, la respuesta elegida al azar puede ser exacta y por lo tanto aumenta la puntuación final; por ello es conveniente que las respuestas incorrectas comporten una sanción, para corregir de esta forma la influencia del azar.

Obviamente también se pueden utilizar otros métodos de corrección; por ejemplo, si se le da importancia al grado de error de las respuestas incorrectas, se puede asignar un punto por cada respuesta correcta y la deducción de un punto por cada error grave y de uno por cada dos errores menores. O simplemente contar los errores y omisiones, deduciéndolos de la puntuación máxima. Estos métodos son menos exactos desde el punto de vista estadístico y metodológico, aunque -como la experiencia muestra- fácilmente aplicables. Sin embargo, es preferible evitar tipologías de puntuación que complican la corrección con el riesgo de distorsiones.

2.6. Evaluación

Completada la corrección de las pruebas se obtiene una puntuación en bruto (*raw score*) que no es significativa, porque varía en función de la duración de la prueba y del tipo de corrección adoptado. Sin embargo, permite una primera revisión completa de la prueba. Se puede hacer una primera lista (por ejemplo, de las puntuaciones más altas a las más

bajas), también se puede obtener la mediana (es decir, la puntuación del examinado que se encuentra a mitad de la lista) o, por último, determinar cuál es la puntuación más común (moda).

Ahora tenemos que convertir el puntaje en bruto en la calificación correspondiente, basada en la escala previamente seleccionada. Por ejemplo, en las universidades italianas la escala es de 0 a 30. En el caso de los exámenes universitarios se debe realizar la conversión de puntaje bruto en treinta puntos. Por lo tanto, a la puntuación máxima alcanzable en la prueba se debe asignar una valoración de 30/30, y las puntuaciones inferiores serán calificadas en proporción. Si, por ejemplo a una puntuación de 50 (máxima puntuación de la prueba) corresponde 30/30, a una puntuación de 40 (es decir, 4/5 del máximo) correspondería 24/30; a una de 30, correspondería 18/30, y así sucesivamente.

Un caso particularmente favorable es un test de treinta *items* con la máxima puntuación de 30. En este caso, la puntuación bruta asignada a la prueba corresponde a la calificación, sin necesidad de conversión.

2.7. Operaciones finales

La evaluación también es completa en este aspecto. No obstante, puede suceder que no sea satisfactoria, porque son demasiados o muy pocos los examinados que alcancen o superen el umbral de promoción (18/30). Nada prohíbe, en el caso de los tests de aprendizaje, que opere una nueva corrección para aumentar o reducir la calificación asignada, sumando o restando a todos los examinados uno o más puntos. No es un procedimiento matemático exacto si bien, en principio, aceptable, ya que deja intacta la lista (para mayor precisión, pero menos práctico, se tendría que multiplicar cada calificación por el mismo coeficiente, por supuesto superior a 1 para aumentar e inferior a 1 para disminuir).

2.8. Verificación

Una vez completados los pasos finales, los tests elaborados se prestan a la realización de algunos estudios con el fin de mejorar las pruebas.

Se pueden analizar los resultados para someter a control el nivel de dificultad o el grado de discriminación de los diferentes *items*; determinar la medida en que cada *item* ha sido resuelto por el grupo que obtuvo la puntuación más alta (cuantificada en el 27%) y la medida en que los han resuelto aquellos que han obtenido la puntuación más baja (también el 27%): si no existe diferencia entre las dos frecuencias o es pequeña, el *item* no discrimina y es prácticamente inútil. Si se comprueba que los errores se reúnen en un determinado *item* o en un pequeño grupo de *items*, el docente también puede obtener una nueva información: se puede tratar (cuando no se deba a la redacción inadecuada del *item*) de insuficiencia de la información ofrecida en el aula, o a una laguna en el texto aconsejado indicado en el programa de la asignatura, etc.

3. Materiales didácticos para las disciplinas de derecho público

3.1. Algunas iniciativas de editoriales italianas

Las Instituciones de Derecho Público, Derecho Constitucional, Derecho Comparado Público (y, especialmente en las Facultades de Ciencias Políticas, de Derecho Constitucional Italiano y Comparado) representan para miles de estudiantes italianos, el primer o por lo menos uno de los primeros contactos con el mundo del Derecho: una aproximación que no resulta siempre fácil, especialmente para aquellos que no pueden asistir con regularidad a las clases.

Las dificultades a las que se enfrenta el neófito del Derecho público son muchas e incluyen: la comprensión de la materia o de ciertas partes de la misma, el primer contacto con el método jurídico, la identificación y clasificación de lo que es esencial saber, y lo que no, los vínculos entre las instituciones y disciplinas ...

Esta situación se ve agravada por las dificultades "logísticas" de no menor importancia: en los grupos cada vez más numerosos, al estudiante casi nunca se le permite resolver sus dudas, durante o después de clase, a través de un diálogo con el profesor; ni puede contactar siempre con su tutor (una figura que a menudo ha quedado puramente teórica) o con otro tipo de colaboradores o asistentes. La verificación de los conocimientos, que precede al examen, se hace generalmente con los sistemas empíricos, confiando en su capacidad para cuestionarse o, como mucho, en la colaboración con compañeros y amigos.

No sólo para realizar la evaluación justa, sino también de estas consideraciones, nace la idea de dar a los estudiantes algunas herramientas pedagógicas que les permitan verificar su nivel de preparación para los exámenes de Derecho Público.

Algunos docentes han elaborado dos volúmenes³ dedicados al Derecho Público y Constitucional italiano y al Derecho Público Comparado (al que se ha añadido más tarde un libro titulado dedicado al Derecho Administrativo⁴), estructurados con la finalidad de ayudar a los estudiantes a verificar su conocimientos, con anterioridad a la realización del examen elaborado siguiendo el modelo del test.

Se dividen (según la edición) en alrededor de mil trescientas o mil quinientas preguntas, agrupadas por temas, en una línea que sigue la tendencia de la mayoría de los manuales. Las secciones en las que se divide cada parte están precedidas por las referencias a las fuentes principales que regulan la materia.

³ L. PEGORARO, A. REPOSO, C. FUSARO, *Istituzioni di diritto pubblico e diritto costituzionale. Itinerari di apprendimento e di verifica*, Giappichelli, Torino, 1995, 2a ed., 1996, 3a ed., 1999, 4a ed., 2002, 5a ed., 2007, e Id., *Diritto comparato pubblico e costituzionale. Itinerari di apprendimento e di verifica*, Giappichelli, Torino, 1998, 2a ed., 2003, 3a ed., 2004.

⁴ L. PEGORARO, L. VANDELLI, R. SCARCIGLIA, V. PARISIO, *Diritto amministrativo. Itinerari di apprendimento e di verifica*, Giappichelli, Torino, 2002.

A las preguntas siguen tres respuestas: una correcta (o la más cercana a la verdad, o la opinión dominante), mientras que las otras dos -si bien posibles- o bien son erróneas, o ilógicas, o menos cercanas a la verdad.

En la parte inferior de cada pregunta, el estudiante se encuentra la referencia a los apartados de los manuales principales en los que se trata el tema (la referencia es a la edición más reciente en el mercado; mientras que en los manuales que por razones de espacio no los mencionan, se puede hacer uso de los índices).

Al lado de cada pregunta aparecen uno, dos o tres símbolos gráficos: son para indicar el grado de dificultad, según los estudiantes que han experimentado hasta ese momento el test. El informe tiene en cuenta, por supuesto, las respuestas alternativas a la más correcta: es evidente que, dadas tres opciones, incluidas preguntas de gran dificultad, es fácil de contestar si las dos respuestas alternativas se encuentran muy alejadas de la correcta⁵.

Las soluciones se encuentran al final de cada parte del libro. Los estudiantes pueden hacer una relectura cuidadosa de las páginas indicadas, comprobando sólo después de la búsqueda -este es el consejo dado por los autores- la exactitud de las respuestas.

La lectura de las preguntas y posibles respuestas permite, por una parte, una nueva revisión del material estudiado en el manual y, en segundo lugar, como se mencionó anteriormente, la comprobación del nivel de preparación.

3.2. Los Tipos de *item* contenidos en los volúmenes

Los volúmenes ofrecen una muestra real de *items*, que también pueden ser utilizados para las pruebas tradicionales, y que permiten un sin fin de combinaciones. Si, por ejemplo, se decide preparar la prueba con treinta temas a elegir entre los disponibles, ¡el número de combinaciones posibles es más de 50!

Las preguntas siguen todas el mismo esquema, pero se han incluido *items* adecuados para determinar los distintos tipos de capacidades.

⁵ Ejemplos de *items* extraídos de L. PEGORARO, A. REPOSO, C. FUSARO, *Diritto comparato pubblico e costituzionale. Itinerari di apprendimento e di verifica*, cit.:

◆◆◆ 198. *En España, la reforma completa de la Constitución necesita de:*

- a) la aprobación de una mayoría de 2/3 de las Cortes, de nuevo la aprobación de las nuevas Cortes con una mayoría de 2/3, el referéndum ()
- b) Aprobación de la mayoría absoluta de las Cortes, referéndum obligatorio ()
- c) la aprobación por mayoría de dos tercios de las Cortes, nueva votación en las Cortes elegidas con aprobación por mayoría absoluta ()

Referencias bibliográficas: de Vergottini Vol. 1, I, III, 10; Morbidelli III, 3; Morbidelli² II, II, 4, 6.

O bien:

◆◆ 1320. *Se define como "abstracto" un modelo de control de la constitucionalidad en el que:*

- a) tiene una importancia fundamental el caso subyacente ()
- b) la norma es considerada por el Tribunal Constitucional en el significado que la jurisprudencia ordinaria le ha asignado ()
- c) la disposición es examinada prescindiendo de su aplicación ()

Referencias bibliográficas: de Vergottini I, III, 17; Morbidelli X, I, 4; Pizzorusso 59; Pegoraro I, 9.

Descartadas las preguntas que requieren respuestas tipo "verdadero-falso" o "correcto-incorrecto" (en este caso la prueba debe ser muy larga para ser fiable), se incluyeron en el texto sólo las preguntas de selección múltiple con una sola respuesta correcta o simplemente con una respuesta mejor; entre éstas, cuestiones del tipo *assertion-reason*, que tienden a verificar la exactitud de dos razonamientos y sus enlaces.

En los volúmenes se encuentran:

a) Preguntas que requieren el mero dominio de las nociones fundamentales, como podría ser, para el Derecho Constitucional español, una pregunta sobre el año de entrada en vigor de la Constitución de Cádiz. A su vez, son incluidas al lado de la respuesta correcta: o dos respuestas incorrectas, pero cercanas a la verdadera (por ejemplo 1815, 1810), o dos respuestas erróneas (por ejemplo 1789, 1878), o, finalmente, una respuesta cercana a la correcta verdad y otra errónea (por ejemplo 1815, 1912)⁶.

b) Preguntas que requieren una mejor comprensión de materia⁷;

c) Preguntas que comportan el ejercicio de la lógica jurídica (por ejemplo, aquellas que imponen combinar categorías similares, como "interpretación doctrinal, jurisprudencial" o "reglamentos de ejecución, los delegados e independientes")⁸.

En función del objetivo perseguido en cada caso, el profesor puede elegir entre un gran número de distintos tipos de preguntas, con el fin de graduar la evaluación de los examinados, o -para usos especiales: por ejemplo, acceso al examen oral- sólo cuestiones de uno u otro tipo.

La aplicación práctica desarrollada durante muchos años ha permitido a los autores comprobar la mayoría de las preguntas, efectuadas con arreglo a los criterios anteriores (claras, inteligibles, concisas, etc.) y -empleando las herramientas ofrecidas por los estudios de docimología que se acaban de mencionar- durante las distintas ediciones,

⁶ Otro ejemplo (extraído de uno de los libros mencionados):

La potestad legislativa que ostentan las Comunidades Autónomas españolas se encuentra limitada por:

- a) la Constitución, el Estatuto, las leyes básicas ()
- b) la Constitución y el Estatuto ()
- c) la Constitución, el Estatuto, las reformas económicas, las leyes del Estado ().

⁷ Ejemplo:

In Spagna, la revisione totale della Costituzione opera con:

- a) approvazione della decisione a maggioranza di 2/3 delle Cortes, riapprovazione delle nuove Cortes con maggioranza di 2/3, referendum ()
- b) approvazione della decisione a maggioranza assoluta delle Cortes, referendum obbligatorio ()
- c) approvazione a maggioranza di 2/3 delle Cortes, riapprovazione delle nuove Cortes a maggioranza assoluta ().

⁸ Cfr. el siguiente:

¿Cuál de estas clasificaciones de la interpretación hace referencia a categorías homogéneas?:

- a) doctrinal, extensiva, literal ()
- b) doctrinal, jurisprudencial, auténtica ()
- c) literal, lógica, extensiva ().

además de ampliar la gama de *items* y actualizar algunos, se han excluido aquellos *items* que no eran suficientemente fiables.

3.3. Disposiciones prácticas para la aplicación de las pruebas

a) Sobre la base de los objetivos perseguidos en cada caso, primero se debe determinar la duración de la prueba, el número de *items*, el tiempo que se asignará. La experiencia muestra que treinta o cuarenta *items* son adecuados, y la duración óptima de la prueba es de unos 30 minutos (es decir, alrededor de un minuto por *item*);

b) A continuación, se deben seleccionar las preguntas, teniendo en cuenta su contenido y grado de dificultad (que se muestra en los libros citados con uno, dos o tres símbolos gráficos, según lo sugerido por los propios estudiantes);

c) los *items* pueden estar dispuestos en el orden que se considere apropiado. Dado que la prueba dispondrá de un plazo determinado, nunca deberían centrarse todos los *items* fáciles al final del cuestionario: la consecuencia sería que los estudiantes más lentos serían penalizados;

d) Para evitar los inconvenientes asociados a la masificación de las aulas durante la prueba, se pueden distribuir los *items* en distinto orden, o dividir a los estudiantes en más aulas y/o trabajar en varios turnos;

e) El reparto de los folios con los *items* no se puede realizar de forma que beneficie a aquellos que los han recibido antes; este reparto debe estar precedido de unas breves instrucciones y de una serie de normas claras que deberán tenerse en cuenta antes, durante y después de la prueba: por ejemplo, «por favor, lea las preguntas atentamente, marque la respuesta en la “hoja de respuestas”», etc.;

f) Una vez recogidas las pruebas, se puede proceder inmediatamente a la corrección y la asignación de puntuaciones a cada *item* según el método elegido, y después se suman para obtener una puntuación total provisional, aplicando las oportunas deducciones si se opta por la modalidad más extendida de corrección;

g) Obtenida la puntuación bruta de cada prueba, hay que traducir estos resultados en una escala de treinta (esta operación no es necesaria si la prueba contiene treinta temas y la puntuación máxima es 30 o, en España, 10);

h) Por último, según el caso, se pueden realizar distintas conversiones, añadiendo o eliminando el mismo número de puntos a todas las pruebas elaboradas.

3.4. Experiencia práctica

Reconociendo los límites sobre la docimología de quien escribe, en el transcurso de más de quince años, al menos se tiene la sensación de que la elaboración de un test bien desarrollado desde una perspectiva científica se puede adaptar a las necesidades prácticas y a la propia gestión de los exámenes universitarios.

Aunque se considera que el criterio correcto para evaluar un test es el que se ha mostrado en el apartado anterior sobre este punto, se prefiere por ejemplo -previo aviso a los examinados- contar las respuestas en blanco y las incorrectas con la misma deducción de un punto, siempre que el propio test esté compuesta de un número adecuado de preguntas. Esto reduce la incidencia de las respuestas aleatorias, pero al mismo tiempo ayuda a reducir los tiempos de corrección, reduce el estrés de los estudiantes y permite el desarrollo del examen oral y/o escrito inmediatamente después de la rápida corrección.

La corrección se ve facilitada por el empleo de papel cuadriculado, y de hecho se completa en unos minutos o unas pocas decenas de minutos.

El resultado de cada candidato se puede obtener -haciendo caso omiso de criterios más complejos- deduciendo un punto de la puntuación máxima, y fijando previamente el umbral de admisión.

Un test que consista en cuarenta *items* parece ser suficientemente selectivo y discriminatorio, y ofrece unos resultados en "campana", con pocas puntuaciones máximas y mínimas, y la mayoría de los resultados concentrados en las franjas intermedias (lo que garantiza la fiabilidad prueba). Por supuesto, como se demuestra científicamente, cuanto mayor sea el número de *items*, más fiable será el resultado, y por esta razón hay quienes han experimentado -como la Universidad de Florencia- pruebas consistentes en un centenar de *items*.

Es posible en casos excepcionales -cuando respecto al promedio del curso muy pocos estudiantes han alcanzado la puntuación mínima exigible, lo que indica que la prueba es posible que fuera demasiado difícil- aumentar el umbral de admisión al examen oral, simplemente incidiendo, partiendo del resultado global, sobre este mínimo (por ejemplo, siendo en Italia 18/30 la calificación que representa el suficiente, nada impide al profesor admitir a la prueba oral o escrita a aquellos que han obtenido la calificación de 17/30 ó 16/30, si advierte que los estudiantes que han obtenido insuficientes representan un número superior al habitual).

Los resultados obtenidos hasta ahora han sido alentadores: los estudiantes aprecian (casi siempre) este método de examen, que asegura una estricta imparcialidad, y se puede combinar con otras herramientas de evaluación, orales o escritas.

No siempre el resultado de las pruebas orales o escritas se corresponde con el resultado del test, si bien esto sucede en la mayoría de los casos. Esta realidad se puede atribuir a las distintas predisposiciones de los estudiantes, su formación, al desarrollo de las capacidades lógicas o al privilegio concedido a la retórica, entre otros.

Sólo por eso, también desde la perspectiva del mercado de trabajo, creemos que una evaluación correcta y completa debe incluir varios tipos de pruebas: Prueba escrita y oral, a la que se pueden añadir evaluaciones conexas, siempre que sea posible, trabajos escritos elaborados por los alumnos e intervenciones durante el curso.

La situación de la Universidad, mortificada por las políticas legislativas y las directrices políticas más cercanas a la cultura mediática de Gran Hermano, no fomenta estas acciones tan duras y exigentes por parte de los profesores, que se encuentran prácticamente obligados, en su gran mayoría, a refugiarse, por razones de tiempo, en las técnicas tradicionales de evaluación, conformistas y poco fiables.

4. Conclusiones

La percepción es que, a pesar de estar tan extendidas, las pruebas tradicionales no son adecuadas para la evaluación del rendimiento de los estudiantes; numerosos docentes de las Facultades de Ciencias Políticas y de Derecho han estado usando los tests para la admisión de los estudiantes al examen oral, haciendo hincapié en la puntuación alcanzada, o simplemente considerando el propio test como barrera para la admisión al oral.

Los procedimientos son, sin embargo, casi siempre aproximados, y la impresión es que el objetivo primordial es simplemente aliviar el volumen de las pruebas tradicionales, el número de candidatos; pero sólo para dar sentido a la prueba oral (o al examen escrito planteado en los mismos términos), dando lugar a todos los problemas expuestos en el apartado primero.

Normalmente los tests están compuestos de un máximo de diez *items*, lo que contribuye a confiar al azar los resultados de muchas de las pruebas, haciéndolas poco fiables; no suele existir un criterio científico para su elaboración; las escasas preguntas son o bien demasiado difíciles, o bien muy similares, o no distinguen entre los distintos conocimientos que se deben tomar como referencia, etc. En resumen, la estructura de la prueba no tiene en cuenta la base científica de la evaluación, de modo que, en efecto, parece que se han elaborado con el único objetivo -la justificación utilitaria- de reducir el número de candidatos admitidos a las pruebas tradicionales que les siguen.

Además, nada se hace para que el estudiante pueda comprobar su preparación sobre la base de este método de evaluación de los conocimientos, y para prepararlo a enfrentarse a las pruebas que, basadas en tests, requiere el mercado laboral: desde las administraciones públicas a las asambleas parlamentarias y regionales, desde los bancos a los tribunales, hasta el sector privado, todos, emplean en la actualidad tests objetivos como el principal criterio de selección de candidatos.

No se puede pedir a los estudiosos del Derecho (y, en nuestro caso, de las disciplinas jurídico-públicas) el conocimiento de la docimología, como en otros casos no se les puede pedir que dispongan de conocimientos profundos sobre otras ciencias⁹, se trata tan solo de evitar aventuras en campos profanos, sin ni siquiera la conciencia de lo que se está haciendo, y la crítica informada, puesto que la evaluación de los estudiantes debe responder a criterios objetivos en lugar de a la improvisación.

⁹ V. L. PEGORARO, *Comparación jurídica y uso “externo” de ciencias “diversas”*, in C. Agurto GONZALES, S. LIDIA QUEQUEJANA MAMANI (coords), *Estudios de derecho civil*, primer volumen, Motivensa editora jurídica, Lima, 2010 (en curso de publicación).