

La *evaluación* de algunas formas de evaluación

The assessment of some forms of evaluation

Resumen

El presente trabajo relata brevemente el origen del examen como instrumento de valoración del aprendizaje. Aborda, asimismo, la necesidad de congruencia epistemológica entre la metodología de la enseñanza, el sustento teórico de un programa de estudios y los instrumentos de evaluación. Finalmente presenta una revisión crítica del llamado *test* de opción múltiple y sus limitaciones.

Palabras clave: Evaluación, racionalidad, *test* estandarizado, coherencia epistemológica, racionalidad crítica, racionalidad técnica

Abstract

This paper presents an overview of the origins of the exam as an instrument to assess learning. It then discusses the need for epistemologic coherence between the didactic method, the theoretical content of an academic program and the methods utilized to evaluate the student. Finally, it offers a critical analysis of the *multiple choice test* and its shortcomings.

Key words: Evaluation, rationality, standardized test, epistemological coherence, critical rationality, practical rationality

Introducción

La aplicación de exámenes es parte crucial en el proceso de evaluación de la enseñanza-aprendizaje en el aula y tiene diversas repercusiones en la vida académica, personal y social de los alumnos que afecta a las instituciones. Este trabajo hace una reflexión sobre el tema de la evaluación con el fin de hurgar en la racionalidad teórica que subyace a las formas de evaluación usadas en el área de las lenguas extranjeras.

Primeramente se hace una breve referencia al origen de la evaluación. Después se revisan dos racionalidades opuestas de las que surgen dos concepciones de evaluación antagónicas. Y finalmente se hace una revisión crítica del llamado test de opción múltiple ampliamente usado para evaluar el aprendizaje de lenguas.

El origen del examen

Existen registros de procedimientos de enseñanza desde épocas antiguas, pero no hay mucha información en relación con la evaluación de los conocimientos adquiridos. Según Judges¹ y Ebel,² un antiguo registro de evaluación formal del que se tiene noticia, se remonta a la China imperial. En esa época se creó el examen como parte de un proceso para elegir a algunos servidores del imperio, provenientes de castas inferiores. Por medio de lo que hoy denominamos exá-

menes de oposición, se seleccionaba a los más talentosos de las castas para integrarlos a la burocracia imperial.

Ebel ubica los exámenes imperiales chinos en el año 2375 a.C. y considera que este sistema de evaluación "constituyó una innovación frente a las formas de organización de las sociedades arcaicas, donde el poder y el privilegio eran prerrogativas hereditarias".³ Se premiaba, explica el teórico, el esfuerzo y el talento de algunos de los súbditos más humildes al permitirles formar parte de la elite gubernamental. Tal vez Ebel no alcanzó a distinguir que pudo haber un tinte demagógico detrás de esta medida, ya que el poder nunca se ha compartido con facilidad.

Antes y durante la Edad Media, entre los siglos v y xv, el examen se introdujo en el ámbito universitario de manera formal, pero sólo podían presentarlo estudiantes sobresalientes, de quienes se esperaba una excelente exposición. De acuerdo con Díaz Barriga,⁴

[...] existe innumerable evidencia de que hasta antes de la Edad Media no existía un sistema de exámenes ligado a la práctica educativa, porque la asignación de notas (calificaciones) al trabajo escolar es una herencia del siglo xx a la pedagogía.

En este sentido, quizá el hecho de ser seleccionado para exponer el talento propio en público, ya fuera en sí una valoración tácita, no formal, del trabajo de excelencia.

En el siglo xvii, Comenio, considerado el precursor de la moderna escuela democrática, incorporó el examen como

¹ Judges, "La evolución de los exámenes". Ángel Díaz Barriga, *El examen: textos para su historia y debate*, pp. 31-48.

² Robert Ebel, "La medición educacional: perspectiva histórica". Ángel Díaz Barriga, *op. cit.*, pp. 83-110.

³ *Ibidem*, p. 86.

⁴ *Ibidem*, p. 13.

parte integral del método de enseñanza-aprendizaje; es decir, lo concibió como parte del proceso pedagógico para apoyar el aprendizaje y lo amalgamó con el método. Exposiciones e intercambios en la práctica diaria daban cuenta de los aprendizajes o bien, de la necesidad de reforzarlos.⁵

Posteriormente, en el siglo XVIII, al crecer la demanda de acceso a la educación surgió la necesidad de medir a los aspirantes en lo individual y de encontrar distintas maneras de selección. Fue entonces cuando las instituciones empezaron a formular normas para la elaboración de exámenes, mismos que solían ser orales.

La asignación de calificaciones al desempeño escolar apareció hasta el siglo XIX, con la formación de sistemas nacionales de educación y para otorgar títulos académicos. Horace Mann concibe un modo de medición por medio de pruebas escritas, aunque sin sustento pedagógico, pues se basaba en los ejercicios cotidianos de clase.⁶

Por su parte, Escudero Escorza señala que en el siglo XX, cobran auge los *test* psicométricos para medir la inteligencia y con ellos, el interés por la medición científica de la conducta humana. Entonces surge la técnica de medición del conocimiento llamada *testing*, misma que se vale de los *test*: pruebas objetivas o estandarizadas para cumplir su función.

Los *test* son resultado de la convergencia de tres corrientes teóricas: a) la positivista, que demandaba la objetividad de la ciencia como la única fuente del conocimiento, y el método científico, como único medio para llegar a él; b) las

teorías evolucionistas de Darwin, que apoyaban la idea de medir las características y diferencias entre los individuos; c) el desarrollo de los métodos estadísticos. La técnica del *test* estuvo fuertemente influenciada por el conductismo, la corriente psicológica de corte positivista imperante en la época.

Según Escudero Escorza,⁷ el desarrollo de la sociedad industrial, que necesitaba de instrumentos de selección para acreditar el conocimiento de los individuos en sus centros de trabajo, fue el contexto idóneo para la introducción del *test* en esta área del quehacer humano. En tanto, Díaz Barriga⁸ asocia la prueba estandarizada con el surgimiento de un proyecto que responde a la necesidad de la industrialización de uniformar a los individuos, quienes son considerados un objeto más entre los procesos de producción. De esta manera, el *test* se erige como el instrumento, presumiblemente, científico para medir la inteligencia, las aptitudes y el aprendizaje, y se implanta en el ejército, la industria y la escuela. Así, la prueba decidía quién era o no apto para ingresar a una de estas instituciones y el lugar que ocuparía.

Durante la segunda Guerra Mundial, esta prueba fue una práctica común en diversas áreas de la vida en Estados Unidos. En esa época, la industria del *test* ya era un negocio floreciente. Se formaron empresas comerciales que proveían estos instrumentos tanto a la iniciativa privada como a las instituciones educativas y gubernamentales. Cabe destacar que desde los años veinte, el *test* ya se aplicaba

⁵ Juan Amós Comenio, *Didáctica magna*, pp. 98-101.

⁶ Robert Ebel, *op. cit.*, pp. 87-88.

⁷ Tomás Escudero Escorza, "From test to current evaluative research".

⁸ Ángel Díaz Barriga, *op. cit.*, pp. 11-28.

en los procedimientos de admisión a universidades. Hoy, la industria de la prueba estandarizada continúa siendo una gran fuente de ingresos tanto en Estados Unidos como en el Reino Unido y muchos otros países.

El fervor por el *testing*, asegura Escudero Escorza,⁹ comenzó a declinar a principios de 1940, pues había movimientos muy críticos respecto al carácter conductista de este recurso. También señala que en la década de los sesenta, en Estados Unidos Ralph Tyler consideraba que la evaluación académica debía sustentarse en una metodología que superara el carácter conductista de las pruebas estandarizadas y que dejara de ser una simple medición, función del *test*. Asimismo, Escudero afirmaba que el propósito más importante de la evaluación debía ser ayudar al alumno a aprender. Para ello propuso incorporar al proceso evaluativo la valoración de diversos componentes significativos en el desempeño del alumno.

Esta rápida reseña del examen, desde la remota China imperial hasta nuestros días, permite apreciar que el examen como forma de evaluación, desde su origen, se vinculó con la vida política y económica de los pueblos, tuvo la función de selección y, en consecuencia, de exclusión. El surgimiento del *test* causó euforia en la sociedad, seguramente por sus supuestas bases científicas y la comodidad de su aplicación. Sin embargo, el paso de la prueba objetiva por el aula escolar ha propiciado situaciones que deben revisarse.

La evaluación escolar y su sustento epistemológico

Para entender el concepto de *evaluación* es necesario tener muy claro su sustento epistemológico, o idea de conocimiento, base del modelo de enseñanza-aprendizaje que subyace a los programas de estudio. Diferentes corrientes filosóficas han dado lugar a distintas percepciones de lo que significa el conocimiento; éstas, a su vez, se han manifestado en concepciones diversas, las cuales dan sustento al enfoque didáctico, al currículo y en consecuencia a la evaluación. Es preciso destacar que tanto prácticas de enseñanza-aprendizaje como formas de evaluación deben coincidir con el concepto de conocimiento en el que se sustentan el enfoque pedagógico y el método de enseñanza. Esta coincidencia redundante en la coherencia epistemológica necesaria entre estos componentes del currículo.

Un ejemplo de congruencia epistemológica se observa en corrientes afines al positivismo, como el conductismo. Para el positivismo, el conocimiento está constituido por hechos y datos empíricos externos al sujeto; se excluyen los procesos mentales del aprendizaje, mismos que son reemplazados por las leyes de la conducta. El conductismo es, según Da Silva y Aline Signoret, una corriente psicológica cuyo objetivo es, precisamente, la descripción, la predicción y el control de la conducta observable, y se ubica en franca afinidad con el positivismo.¹⁰

Al respecto, Juan Manuel Álvarez resalta la concepción simplista del conductismo por el hecho de dejar fuera de

⁹ Tomás Escudero Escorza, *op. cit.*

¹⁰ Helena Da Silva y Aline Signoret, *Temas sobre la adquisición de una segunda lengua*, p. 64.

su perspectiva el carácter exploratorio del proceso de enseñanza-aprendizaje.

Del positivismo surge la pedagogía por objetivos (interpretación y aplicación lineal del conductismo al currículum) que reduce el conocimiento a una lista de objetivos empíricamente observable. El currículum se resume en una serie de programaciones en torno a objetivos. Con ellas se trata de controlar, de modo experimental y externo al sujeto que aprende, el desarrollo del mismo y predecir los resultados separando medio de fines. Al definir los objetivos en términos de una lista de realizaciones concretas (el aprendizaje reducido a conductas observables) se inhibe la naturaleza exploradora de los procesos de la enseñanza y del aprendizaje.¹¹

El llamado enfoque estructuralista en la enseñanza de lenguas extranjeras ejemplifica la congruencia epistemológica entre el positivismo, el enfoque pedagógico, el método y las prácticas de evaluación. Para el estructuralismo, el aprendizaje de una lengua extranjera supone la adquisición de un nuevo conjunto de estructuras lingüísticas, nuevas conductas, que deben afianzarse mediante el refuerzo. En este sentido, para proporcionar al alumno el refuerzo requerido se crearon los *drills*, ejercicios repetitivos intensos que, se creía, permitirían la adquisición de las estructuras de la lengua extranjera de forma automática, sin tener que pensar en ellas; es decir, se suponía que el conocimiento se lograría con base en la repetición por medio de *drills*. De acuerdo

con este enfoque –muy popular hasta los años setenta–, el conocimiento es un acondicionamiento conductista, no demanda razonamiento alguno por parte del aprendiente, y la evaluación, en general, se efectúa por medio de *test* objetivos.¹²

Si en su práctica el educador se limita a transmitir datos y hechos sin propiciar la reflexión y la discusión acerca de lo que significan, entonces la idea de conocimiento es congruente con la corriente positivista del aprendizaje. De esta manera, es posible afirmar que este procedimiento, con algunos cambios y con nombres “modernos”, prevalece en la enseñanza de lenguas extranjeras.

La evaluación del aprendizaje en el ámbito escolar desde dos racionalidades opuestas

Se mencionó antes que la manera de evaluar los conocimientos se vincula, necesariamente, a la forma como el profesor conduce su enseñanza; o sea, la racionalidad de un sistema de enseñanza determina su forma de evaluación. Se entiende por racionalidad, el conjunto de procedimientos que subyacen al proceso de enseñanza-aprendizaje:

[...] el conjunto de supuestos y creencias, así como de prácticas, del profesor sobre la enseñanza y el aprendizaje, sobre la escuela, el alumno, el conocimiento [...]

¹¹ Juan Manuel Álvarez Méndez, *Evaluar para conocer, examinar para excluir*, p. 28.

¹² Parkinson de Saz, “El estructuralismo y la enseñanza de las lenguas”. María del Carmen Koleff, *Técnicas metodológicas y observación de clase*, pp. 215-234.

que dan lugar a unas prácticas evaluativas y no a otras.¹³

La evaluación se puede conceptualizar desde dos racionalidades diferentes y antagónicas. La primera es la racionalidad técnica, también llamada *instrumentalista*, en la cual el profesor suele ser un agente externo en el proceso de evaluación. La segunda es la perspectiva denominada *racionalidad crítica*, según la cual la evaluación es una actividad crítica del aprendizaje y, en sí misma, es aprendizaje.

Desde la perspectiva positivista, en la que abrevia la racionalidad técnica, la conducta externa, observable, es el dato empírico que evidencia el aprendizaje; por lo tanto, la enseñanza consiste en modificar la conducta del aprendiente, sin ninguna intención de incentivar su reflexión ni su capacidad crítica, tal como se hace con la corriente estructuralista en la enseñanza de las lenguas.

La racionalidad técnica o instrumental se asocia con la pedagogía por objetivos. Según su postulado básico, el desarrollo del conocimiento sobre los fenómenos educativos es resultado de la observación y medición de los hechos. Para efectos de evaluación, las implicaciones surgidas de la racionalidad técnica son: el juicio sobre el fenómeno educativo, el cual se hace desde lo observable y medible, debe ser objetivo y neutral y, entonces, la unidad de análisis se plantea sobre la conducta observable; su influencia positivista resulta evidente.

Desde la racionalidad técnica, las pruebas estandarizadas u objetivas son el recurso idóneo que, se cree, permite

eliminar los posibles sesgos subjetivos de la evaluación aplicada por el profesor. De este modo, la evaluación no toma en cuenta los procesos mentales del conocimiento y elimina todo lo que pueda ser subjetivo, para dar paso a la objetividad. El conocimiento se asume libre de valores y neutral. La evaluación es sumativa; es decir, se limita a medir el conocimiento según estos principios y a emitir una calificación. Por lo general, la participación del profesor se reduce a la aplicación de técnicas de evaluación; el maestro es un aplicador del examen que le proporciona un grupo de colegas o una compañía comercial encargada de elaborarlos. La racionalidad técnica es el sustento epistemológico del *test* o prueba objetiva.

Sería deseable que la evaluación, elaborada desde los principios mencionados, fuera el medio ideal para valorar el aprendizaje; pero la práctica docente cotidiana lo pone en duda: no siempre es posible formular objetivos de enseñanza en términos de conductas observables; no siempre se pueden aislar los contenidos de su contexto. Intentar forzar la situación suele llevar a trivializar, simplificar o distorsionar los contenidos de aprendizaje que se requiere evaluar.

Al respecto, se podría afirmar que no todo conocimiento puede valorarse por medio de objetivos conductuales. Para Salinas, por ejemplo, una implicación adversa del modelo de objetivos reside en:

[...] limitarnos el objeto de la educación al objetivo prefijado con antelación en forma de conducta observable. Entonces, la evaluación puede ser identificada con un proceso fundamentalmente técnico de comprobación o medición del alcan-

¹³ Dino Salinas, *¡Mañana examen! La evaluación: entre la teoría y la realidad*, p. 35.

ce de las conductas definidas en forma de objetivos.¹⁴

De lo anterior se desprende que el modelo de evaluación por medio de *tests* o pruebas objetivas es congruente con la racionalidad técnica que concibe al alumno como un mero *receptor de información*. Y como se dijo anteriormente, esta racionalidad técnica ocupa un lugar preponderante en la evaluación del aprendizaje hoy en día.

La racionalidad crítica y la evaluación

Desde concepciones epistemológicas distintas y opuestas respecto del positivismo, de la hermenéutica, de la nueva sociología, de la teoría crítica, de la epistemología genética y del constructivismo (psicológico y social) surgen concepciones del conocimiento como construcción histórica y social, dinámica que necesita de contexto para ser entendido e interpretado. Aquí, el currículum se entiende como construcción histórica y sociocultural.¹⁵

La perspectiva crítica en la educación implica el involucramiento activo, participativo y crítico, tanto del profesor como del alumno, en la búsqueda del conocimiento. El profesor incentiva la curiosidad por aprender y apoya al alumno para que explore y le dé significado a esos aprendizajes:

La tarea de la educación que surge de esta visión dinámica del conocimiento es ayudar a quien aprende a desarrollar

reflexivamente un conjunto de modos de pensamiento o modos de aprendizaje de contenidos que son considerados valiosos en la sociedad.¹⁶

La educación, desde el punto de vista de la *racionalidad crítica*, no se limita a transmitir información, sino que persigue alentar la exploración y la reflexión en torno a los contenidos de aprendizaje. Por lo tanto, la valoración educativa, necesariamente, debe ser sustento crítico y formar parte integral del proceso de aprender. La evaluación de tipo crítico da cabida a la confrontación de argumentos entre profesor y alumnos, frente a los resultados del examen; propicia el intercambio de ideas y puntos de vista, y contribuye así, a la formación del individuo; condición ineludible en el ámbito de la educación.

En el mismo tenor, Álvarez considera que, de acuerdo con las tendencias actuales, si el profesor entiende la educación como acceso a la cultura y al saber, y como patrimonio universal, su reto consiste en hacer de la evaluación un recurso más de aprendizaje. Desde esta perspectiva de la educación, la racionalidad crítica proporciona el sustento teórico al currículum, al proceso de enseñanza-aprendizaje y a la evaluación.

Por su parte, Salinas considera que programar la enseñanza por objetivos puede tener alguna utilidad, siempre y cuando los objetivos respondan a un proyecto académico y social que permita planear actividades significativas. No obstante, se corre el riesgo de que al tratar de dividir los objetivos en unidades mínimas de

¹⁴ *Ibidem*, p. 42.

¹⁵ Juan Manuel Álvarez Méndez, *op. cit.*, p. 28.

¹⁶ *Ibidem*, p. 30.

EVALUACIÓN DESDE LA RACIONALIDAD CRÍTICA	EVALUACIÓN DESDE LA RACIONALIDAD TÉCNICA
Acción comunicativa	Acción estratégica
Formativa	Sumativa
Referida a principios educativos	Referida a criterios o criterial, y a normas o normativa
Pruebas de ensayo, de elaboración y de aplicación	Exámenes tradicionales: pruebas objetivas
Evaluación del aprendizaje	Medida del rendimiento escolar
Enseñanza dirigida a la comprensión	Enseñanza dirigida al examen
Lo valioso se identifica por su interés formativo	Lo valioso se identifica con lo que más puntúa
Implicación-compromiso del profesor	Distanciamiento en nombre de la imparcialidad
Actividad de aprendizaje	Acto de control y sanción

conducta, se trivialice lo que se pretende evaluar.¹⁷

Según los dos enfoques mencionados, las características de la evaluación se resumen en el siguiente cuadro.¹⁸ Se observa que la racionalidad crítica se encamina más a las corrientes pedagógicas progresistas, que promueven el aprendizaje como resultado de la actividad mental intensa tanto del alumno como del maestro. Por el contrario, la racionalidad técnica se ciñe a experiencias tradicionalistas que se basan en actitudes pasivas, en las cuales el alumno es receptor y repetidor de información.

El test de opción múltiple: una revisión crítica, alcances y limitaciones

El test objetivo o estandarizado está presente, y en un nivel mundial, en las prácticas de evaluación escolar y universitaria. Es un hecho que en Estados Unidos, las asignaciones monetarias para la educación pública dependen, en buena medida, de los resultados que las poblaciones estudiantiles obtienen en este tipo de pruebas. En los países pertenecientes a la Organización para la Cooperación y el Desarrollo Económico (OCDE), las pruebas llamadas Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) y Programme for International

¹⁷ Dino Salinas, *op. cit.*, p. 46.

¹⁸ Juan Manuel Álvarez Méndez, *op. cit.*, pp. 20-22.

Student Assessment (PISA), se valen de *test* estandarizados para medir el rendimiento escolar, internacionalmente; esto sugiere las dimensiones del alcance de dichas pruebas.

Muy conocida es la industria del *test* para medir la lengua inglesa como lengua extranjera. Al respecto, Alderson¹⁹ refiere en sus trabajos algunas de estas grandes empresas. Entre los *test* más conocidos se cuentan: TOEFL, Test of English as a foreign language y la Prueba del inglés como lengua extranjera; son utilizados para certificar el dominio (*proficiency*) de la lengua inglesa y, año con año, se aplican a miles de estudiantes de inglés en el mundo. Asimismo, existe una amplia gama de este tipo de *test*, para diferentes propósitos y niveles de conocimiento en inglés.²⁰

Diversos investigadores han cuestionado la posibilidad de este instrumento para evaluar por sí solo, de manera "objetiva y válida", el rendimiento escolar. Los teóricos coinciden en que no hay manera de que este tipo de *test* dé cuenta de los procesos cognitivos tan complejos, implicados en la actividad de aprender. Se acepta que sirve para evaluar la adquisición de información en sus manifestaciones más simples, como la memorización de datos o las habilidades de bajo nivel; no obstante, no puede ser la base de la decisión para determinar si el aprendizaje se lleva a cabo o no, en situaciones en que se requiere el uso de habilidades de niveles superiores.

Desde hace algunas décadas, se ha insistido en la necesidad de evitar el apren-

dizaje memorístico y se ha pugnado por instituciones con modelos educativos de corte crítico, que formen al aprendiente como individuo pensante, reflexivo y crítico, capaz de emitir opiniones y de aplicar los conocimientos adquiridos en la escuela en diferentes ámbitos de su vida. Estos principios han sido impulsados desde la Secretaría de Educación Pública (SEP), en México, en coordinación con la UNESCO.²¹

En los trabajos de estas instituciones, elaborados por académicos destacados y difundidos en todo el mundo, se percibe la influencia de investigadores con tendencia a la racionalidad crítica. Si esta corriente se está impulsando desde la escuela primaria, los enfoques críticos deberían ser el sustento epistemológico-pedagógico y, en consecuencia, de la evaluación. Pero el *test* como instrumento de medición rompe con los principios de estas corrientes de pensamiento, al limitar las posibilidades de expresión del aprendiente a formatos de examen que demandan respuestas unívocas e impuestas como verdades únicas.

Esto sugiere la incongruencia epistemológica ya mencionada: por un lado, se habla de la educación como el medio para formar estudiantes críticos, analíticos y pensantes; por el otro, se niega la posibilidad de aplicar estas capacidades en los exámenes. De esta manera, se presenta una forma de evaluar inconsistente, con los principios que rigen actualmente y la racionalidad con que se concibe el aprendizaje, avalados

¹⁹ Charles Alderson, C. Clapham y D. Wall, *Exámenes de idiomas. Elaboración y evaluación*, pp. 249-250.

²⁰ Categorización de Alderson. *Ibidem*.

²¹ Secretaría de Educación Pública, *Plan y Programas de Estudio*, 1993; Jacques Delors, *Informe Delors. La educación encierra un tesoro*; Edgar Morin, *Los siete saberes necesarios para la educación del futuro*.

por instituciones como la UNESCO e investigadores progresistas.

Es preciso enfatizar, entonces, que esta flagrante e inaceptable incongruencia debe ser discutida por los cuerpos docentes. Sin duda sorprenderá encontrar que, aun cuando algunos programas anuncian adherirse a principios de sustento crítico, sus técnicas, materiales didácticos y de evaluación se mantienen en la racionalidad técnica.

La técnica del test de opción múltiple desde una perspectiva crítica

Díaz Barriga observa serias limitaciones de tipo epistemológico en el *test objetivo o estandarizado*. Sostiene que no existe una teoría de evaluación en la cual se sustente el *test* objetivo, por lo que no puede ser un instrumento confiable, supuestamente característica distintiva de estas pruebas. Más adelante menciona tres de los pasos técnicos de la construcción del *test*:

1. Atributo: se refiere a lo que se mide, inteligencia o aprendizaje.
2. Operacionalización del atributo: es la forma como se hace la medición.
3. Construcción de instrumentos: es la base epistemológica sobre la cual se construye el *test*.

El investigador señala que entre estos pasos hay un vacío epistemológico insalvable: "entre la asignación de un atributo y la asignación de operaciones para reconocerlo existe una independencia total que no está mediada por una construcción conceptual que resulte compren-

siva del fenómeno".²² Para el teórico de la educación, este hecho deja totalmente al criterio de quien elabora el *test* la asignación de las operaciones que, a su juicio y sólo a su juicio, son válidas para reconocer la presencia del atributo.

El siguiente ejemplo referido por Hoffmann²³ ilustra las consideraciones de Díaz Barriga. El 18 de marzo de 1953, el diario *Times* de Londres publicó la carta que un lector había dirigido al editor del periódico. En ella narraba que su hijo de once años había presentado un examen de admisión escolar, el cual incluía la pregunta sobre deportes nacionales en el Reino Unido: *Choose the odd one out among* (De las cuatro opciones, selecciona una que es diferente): entre las opciones estaban cuatro deportes arraigados en el Reino Unido: *cricket, football, billiards y hockey*.

El lector indica haber elegido *billiards* porque es el que se juega intramuros; un colega suyo seleccionó *football* porque es el único en el cual la pelota no se golpea con un implemento; un vecino dijo que es el *cricket* porque sólo este deporte no tiene como objetivo poner la pelota en una red; su hijo –quien presentó el examen– dijo que el *hockey* porque es el que juegan mujeres, según su experiencia en los campamentos de verano.

El 19 de marzo, el *Times* imprimió dos cartas más con opiniones de dos lectores: uno decía que la correcta era *billiards* porque no es juego de equipo; el otro, que la correcta era el *football* porque para jugarlo se usaba una pelota inflada y no sólida, como en los otros deportes. Al día siguiente, un profesor

²² Ángel Díaz Barriga, *op. cit.*, p. 20.

²³ Banesh Hoffmann, *The Tyranny of Testing*, pp. 17-21.

reportó los siguientes resultados, obtenidos en un grupo que había tomado el mismo *test*: *football* 18, *billiards* 17, *hockey* 3, *cricket* 1. Las respuestas y sus justificaciones siguieron llegando a la redacción, hasta que el periódico decidió poner fin a la discusión.

El ejemplo ilustra cómo, ante la falta de una teoría de la evaluación que sustente la elaboración de los reactivos del *test*, todas las opciones pueden sostenerse con tantos argumentos como hay lectores; pero, arbitrariamente, sólo hay una respuesta: la que tuvo en mente quien elaboró el *test*. Muchos ejemplos de este tipo se multiplican en los *tests objetivos*, aun en aquellos elaborados por equipos de expertos que cuentan con asesores especialistas, en las diferentes materias objeto de evaluación.

Díaz Barriga²⁴ plantea otras objeciones a la evaluación, con base en la prueba objetiva. Observa que es el debate de los problemas de objetividad, validez y confiabilidad de la prueba, lo que le ha dado el matiz cientificista; sin embargo, afirma que estas discusiones son meramente técnicas y se dirigen a problemas como la elaboración, tipo de *tests* y su validación estadística, pero ignoran un aspecto central: el problema pedagógico del aprendizaje. Al centrar la atención en el examen, se desvirtúa el aprendizaje.

Los procesos para elaborar estas pruebas, como especificaciones para la redacción de reactivos, procedimientos de ensayo y error (*piloteo*), revisión, validación y demás parafernalia que las acompaña,²⁵ ocupan tiempo, esfuerzo y

recursos inmensos. Estos recursos bien podrían canalizarse a la actividad pedagógica, pero en la práctica, desafortunadamente, ésta queda relegada a un segundo plano. Un ejemplo de ello: llama la atención la clasificación que Munby²⁶ hace de doce áreas comunes de error en "distractores" u opciones de los reactivos de *test* de comprensión de lectura en inglés; tres de estas áreas de error son: reacción estética fuera de lugar (haber sido atraído por una palabra o frase llamativa), malinterpretar el tono emocional del texto y la incapacidad para entender el lenguaje figurado. Con estos tres aspectos es posible imaginar el tiempo requerido para analizar los errores, en decenas o tal vez cientos de reactivos, y preguntarse si ese tiempo y los recursos pudieron usarse en el tema central: el problema del aprendizaje; además, salta a la vista la poca factibilidad de un sustento teórico que dé validez a esta clasificación del autor.

Hoffmann²⁷ ofrece una revisión acuciosa del *test* y coincide con Díaz Barriga en que alrededor de los temas de la objetividad, la confiabilidad, la validez y otros términos "científicos" empleados en la elaboración de los reactivos, hay mitos que no se sostienen. Por ejemplo, Hoffman afirma que la supuesta objetividad de la prueba no reside en ésta, sino únicamente en el hecho de que ningún elemento de subjetividad entra en el proceso de calificación, una vez que se decidió la clave de respuestas correctas. Es de esperarse, entonces, que cada reactivo y sus distractores estén influidos

²⁴ Ángel Díaz Barriga, *op. cit.*, p. 20.

²⁵ Charles Alderson et al., *op. cit.*, pp. 14-41.

²⁶ Charles Alderson, *op. cit.*, pp. 204-205.

²⁷ Banesh Hoffmann, *op. cit.*, pp. 17-21.

por la subjetividad del redactor, quien no necesariamente es maestro, quizá sea un técnico elaborador de pruebas objetivas.

También observa que cuando se redactan reactivos para un *test*, y se intenta que éstos *discriminen*, esto es, que cumplan la función de *distracer o trampear* al examinado, se procede a buscar palabras que lo alejen de la respuesta correcta, pero que parezcan *plausibles*. Entonces surgen respuestas plausibles y que pudieran ser la mejor, como suelen pedir las instrucciones. Esta búsqueda de alejar a las plausibles de la mejor respuesta propicia “crear dificultad espuria por medio de introducir ambigüedad”.²⁸

It is not without significance that the professional testers refer to “wrong” answers as “distractors”, “misleads”, and “decoys”. The decoys are deliberately designed to seem plausible. They are, in fact, deliberate traps. Were they always traps baited with definitely spurious bait one might tolerate them, even though their presence gives the test an air of trickery and deception that is not altogether becoming. But often the traps are baited unfairly. For it is difficult to draw a sharp line between legitimate wile and illegitimate deceit.²⁹

²⁸*Ibidem*, p. 69.

²⁹*Ibidem*, pp. 69-70. “No es casual que los elaboradores profesionales les llamen a las opciones no correctas ‘distractores’, ‘engaños’ o ‘señuelos’. Los distractores son trampas deliberadamente diseñadas para parecer plausibles. De hecho, son trampas planeadas con la intención de hacerlas espurias, tal vez se podrían tolerar, aunque su presencia le daría al test un aire fraudulento y de dolo que no le va bien. Pero con frecuencia, las trampas son injustamente amañadas. Entonces es difícil trazar la línea entre la persuasión legiti-

Para Hoffmann, el propio formato del *test* compele a la ambigüedad, pues lograr reactivos genuinamente adecuados resulta poco factible. Por ejemplo, para elevar el nivel de dificultad de los *distractores* que las estadísticas calificaron de “muy fáciles”, suele incluirse en las opciones términos poco conocidos, sofisticados o seleccionados de manera artificial y, en este punto, es prácticamente imposible no caer en ambigüedades. Por lo tanto, es evidente que se pone en duda la objetividad del instrumento.

En otro aspecto, el mismo autor se pregunta y reflexiona: por qué, en general, los *test* piden que se seleccione “la mejor respuesta” o “la que va mejor” al contexto, y no solicitan, llanamente, la “respuesta correcta”. Este hecho es ya una aceptación de posible ambigüedad del reactivo. ¿Qué criterio utilizar para decidir cuál es “la mejor”? ¿Qué debe hacer el alumno, quien tiene argumentos sólidos para defender otra respuesta, que considera correcta? Si la validación estadística decidió que la respuesta correcta es sólo una, entonces los papeles se invierten y el *test* sirve a las estadísticas, no al proceso de aprendizaje. En estos casos, pareciera que se penaliza al alumno que más sabe y puede ver desde diferentes ángulos, el problema planteado. Por esta razón, Hoffmann afirma que el *test* es una camisa de fuerza para el estudiante mejor dotado.³⁰

Los redactores de pruebas objetivas argumentan que los reactivos ambiguos se analizan para encontrar, estadísticamente, indicadores de ambigüedad y eli-

ma y el engaño ilegítimo.” (Ésta y las siguientes traducciones son nuestras.)

³⁰*Ibidem*, p. 92.

minarlos; pero estos resultados tienen valor limitado, pues al no haber una teoría que sustente la elaboración de distractores, la corrección puede ser “vuelta a empezar”.

Por otra parte, para Álvarez Méndez el examen, en cualquiera de sus formas, se puede utilizar como instrumento para aprender o sólo para calificar. En el segundo caso lo que se mide es únicamente si los alumnos son “buenos receptores y buenos memorizadores pero no si son críticos o creadores o si son capaces de aplicar los conocimientos que adquieren”. En esencia, ésta es la función del *test* objetivo.

Uno de los problemas del examen que exige respuestas atomizadas es que “obliga” a repetir las mismas vulgaridades o vaguedades previamente contadas, sobre el pensamiento o conocimiento transmitido. Tal vez éste sea importante pero permanece oculto, indiscifrable o confundido entre tanta palabra vacía que simplifica el contenido hasta hacerlo incomprensible y, consecuentemente, hasta hacerlo reproducible en los mismos términos en que fue contado, sin salirse una línea, una palabra del guión dado con anterioridad.³¹

De acuerdo con lo anterior, resulta evidente que el *test* de opción múltiple no puede ser un instrumento adecuado para evaluar el proceso de aprendizaje, en toda su complejidad. No obstante, en la práctica se acepta que una prueba objetiva, diseñada con mucho cuidado, se aplique en ciertas ocasiones para averiguar, por

ejemplo, si se retuvo algún dato o fecha importante.

Estas técnicas, según Álvarez Méndez, no

[...] informan de qué modo las respuestas que se pueden dar a las preguntas que se formulan, pueden representar correctamente la experiencia de aprendizaje y el proceso de pensamiento que lleva a expresar conocimiento.³²

Hoffmann³³ resume de la siguiente manera las principales críticas al examen de opción múltiple, recogidas de importantes investigadores:

- Niegan a la persona creativa la oportunidad de mostrar su creatividad, y favorecen al dócil y al astuto por encima de quien tiene una opinión propia.
- Penalizan al candidato que percibe aspectos sutiles no registrados por gente menos capaz, ni siquiera por los propios redactores del *test*.
- Tienden a ser superficiales e intelectualmente deshonestos, con preguntas difíciles, hechas artificialmente, ya que preguntas genuinas y de contenido profundo no tienen cabida en el formato de la opción múltiple.
- Sólo toman en cuenta la elección de la respuesta y no la calidad de pensamiento que condujo a tal elección.

³¹ Juan Manuel Álvarez Méndez, *op. cit.*, p. 96.

³² *Ibidem*, p. 91.

³³ Banesh Hoffmann, *op. cit.*, p. 150.

- Con frecuencia se convierten en juegos subjetivos de adivinanzas, que poco tienen que ver con el aprendizaje.

Aunque se mencionó antes que el examen de opción múltiple puede ser útil para medir el aprendizaje de datos precisos, es obvio que sus limitaciones exceden sus bondades.

La lista de Hoffmann deberá motivar la reflexión sobre el uso de la prueba objetiva o estandarizada. Sin duda, la poderosa industria del *test* ha deslumbrado tanto a los administradores escolares como a los profesores, con el argumento de que tales instrumentos tienen el respaldo de la ciencia de la estadística. La realidad es que su debilidad supera su fortaleza. Por fortuna hay investigadores críticos que están aportando luz al tema, y es posible informarse de sus alcances y limitaciones.

Muestra de ello es Charles Alderson, investigador reconocido en el área de la evaluación, quien ha observado cambios en relación con la evaluación de la lectura en lengua inglesa, adelanta su posible decadencia:

Recent years have seen an increase in the number of different techniques used for testing reading. Where multiple-choice prevailed, now we see a range of different "objective" techniques, and also increase in "non-objective" methods, like short-answers questions, or even the use of summaries which have to be subjectively evaluated. Test constructors often have to use objective technique for practical reasons, but there is a tendency for multiple-choice to be avoided if at all possible although the use of computer-

based testing has resulted in a, hopefully only temporary, resurgence of multiple-choice techniques.³⁴

Hoffman, por su parte, sostiene que no existe ni existirá un solo instrumento de evaluación que pueda dar cuenta del aprendizaje humano. Reitera que la complejidad de las habilidades cognitivas es enorme, interactiva y multifacética; asimismo, pone en duda que los instrumentos disponibles sean los idóneos. Sin embargo, acepta que la evaluación es parte de la vida de las sociedades y que es necesario seguir investigando al respecto. No es posible seguir privilegiando la comodidad de la aplicación de la prueba objetiva, en detrimento del desarrollo intelectual de los alumnos y de la educación.

Discusión

Desde la antigüedad, la evaluación ha sido parte de la vida en las sociedades. Comenio, el gran educador, concibió la evaluación como parte del proceso de enseñanza-aprendizaje. Es de lamentar

³⁴ "En años recientes se ha visto un aumento en el número de técnicas usadas para examinar la lectura. Ahora vemos una variedad de técnicas 'objetivas', donde antes prevalecía la opción múltiple; también ha habido un aumento de métodos 'no objetivos', como las preguntas de respuesta corta o hasta resúmenes que deben ser evaluados de manera subjetiva. Los elaboradores de tests, a menudo tienen que usar la técnica objetiva por razones prácticas, pero la tendencia es a evitar la opción múltiple, si es posible del todo (aunque examinar por medio de la computadora ha resultado en un resurgimiento de las técnicas de opción múltiple, que se espera sea sólo temporal)." Charles Alderson, *Assessing Reading*, p. 205.

que esta orientación se haya desvirtuado con el paso del tiempo.

El *test* objetivo surge con la expansión de la industria, por la necesidad de contar con un medio confiable para detectar las habilidades y capacidades de los aspirantes a ingresar al ámbito laboral. La idea, ampliamente divulgada, de que el *test* era una forma científica y objetiva de evaluación, lo convirtió en una herramienta muy popular en el mundo; esta percepción influyó para que se adoptara en instituciones educativas de todos los niveles.

En el mismo sentido, se ha llegado a la conclusión de que la técnica de elaboración del *test* objetivo es endeble, pues no tiene un sustento teórico que la apoye. Incluso los expertos en redactar este tipo de prueba señalan la dificultad de elaborar reactivos (*ítems*) confiables y válidos; de esta manera, ante la falta de una teoría que les dé sustento, las sugerencias para corregir las fallas caen inevitablemente en simples recomendaciones de buena fe.

Propuestas

Es necesaria la búsqueda colegiada de instrumentos de evaluación consistentes con la filosofía de la enseñanza-aprendizaje, propuesta por la institución educativa. Se requiere determinar cuál es la racionalidad o visión epistemológica que subyace a los procesos de enseñanza-aprendizaje y decidir si la prueba objetiva contribuye en la formación prevista para el egresado universitario. De no ser así, la evaluación seguirá siendo únicamente un instrumento para dar o negar a los alumnos, el pase al curso inmediato

superior, y dejará pasar el potencial formativo de las propuestas críticas para la valoración del conocimiento.

Finalmente, el presente trabajo no puede pasar por alto el hecho de que el conocimiento de lenguas extranjeras es una herramienta muy valiosa e indispensable para la investigación, la comunicación y el acercamiento a la cultura universal.³⁵ Es necesario, asimismo, revisar si los procedimientos de enseñanza y prácticas de evaluación actuales son las idóneas para la óptima formación de los universitarios.

Bibliografía

- Acuerdo Nacional para la Modernización de la Educación Básica*, México, 1992.
- Alderson, Charles. *Assessing Reading*. Cambridge, Cambridge University Press, 2010.
- Alderson, Charles, C. Clapham y D. Wall. *Exámenes de idiomas. Elaboración y evaluación*, Cambridge, Cambridge University Press, 1998.
- Álvarez Méndez, Juan Manuel. *Evaluar para conocer, examinar para excluir*. Madrid, Ediciones Morata, 2001.
- Comenio, Juan Amós. *Didáctica magna*. México, Porrúa, 1970.
- Da Silva, Helena y Aline Signoret Doscarberro. *Temas sobre la adquisición de*

³⁵ Por ello, la Universidad Autónoma Metropolitana debe ser una institución de vanguardia en este campo, como lo es en muchos otros, y sus egresados tienen que poseer una formación sólida en cuanto a lenguas extranjeras, que les permita enfrentar con éxito los retos de su futura vida académica y laboral.

- una segunda lengua*. México, Universidad Nacional Autónoma de México-Centro de Lenguas Extranjeras, 1996.
- Delors, Jacques. *Informe Delors. La educación encierra un tesoro*. París, UNESCO, 1996.
- Díaz Barriga, Ángel. *El examen: textos para su historia y debate*. México, Universidad Nacional Autónoma de México-Centro de Estudios sobre la Universidad/ Plaza y Valdés, 2001.
- Hoffmann, Banesh. *The Tyranny of Testing*. Nueva York, Dover Publications, 1962.
- Koleff, María del Carmen. *Comp. Técnicas metodológicas y observación de clase*. México, Universidad Nacional Autónoma de México-Centro de Estudios para Extranjeros, s/f.
- Morin, Edgar. *Los siete saberes necesarios para la educación del futuro*. París, UNESCO, 1999.
- Salinas, Dino. *¡Mañana examen! La evaluación: entre la teoría y la realidad*. Barcelona, Grao, 2001.
- Secretaría de Educación Pública. *Plan y Programas de Estudio*. México, 1993.

Cibergrafía

- Escudero Escorza, Tomás. "From tests to current evaluative research. One century, the 20th. of intense development of evaluation in education".
http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1eng.htm (consultado en diciembre de 2012)