

INVESTIGACIÓN

La compleja elaboración de exámenes, 16 razones para utilizar la opción "no sé"

*Felipe Tirado**
*Eduardo Backhoff***

Resumen:

Prácticamente en cualquier institución educativa se construyen pruebas de evaluación, sin embargo ésta es una tarea compleja que pocas veces se realiza con las debidas consideraciones. El presente artículo hace una serie de reflexiones en torno a la elaboración de pruebas de opción múltiple, centrándose en la relevancia de incluir una alternativa para que, cuando se ignora la respuesta, se exprese "no sé".

En el artículo se ofrecen, además de argumentos educativos a favor del uso de la opción "no sé", una serie de planteamientos basados en investigación sistemática, al analizar estadísticamente la evidenció empírica encontrada, la cual agrega razones importantes y favorables para recomendar el uso de esta opción. Lo más relevante de este trabajo es que los resultados de la investigación que se presenta apuntan a favor del procedimiento propuesto, ya que permite mejorar la validez criterial, concurrente y predictiva de un examen, lo que es muy relevante en la elaboración de los mismos.

Abstract:

Evaluation tests are employed in almost all educational institutions. However this is a complex task which is rarely given due consideration. This article includes a series of reflections on the design of multiple choice tests. It focuses on the relevance of including the option "I don't know" as a viable reply when the answer is unknown.

The article offers educational arguments in favor of the "I don't know" option and also a series of proposals based on systematic research, through a statistical analysis of the empirical evidence found. These provide favorable reasons for recommending the use of this option. The most relevant aspect of this study is that the results of the research presented favor the proposed procedure, since it improves the concurrent and predictive criterion validity of an exam, which is relevant to exam design.

Introducción

En toda empresa se tiene un propósito y se busca, en principio, el cabal cumplimiento de éste; de aquí que resulta necesario, o aún más, sustantivo, poder evaluar el grado con que se ha logrado lo propuesto.

* *Director de la ENEP Iztacala UNAM. CE: tirado@servidor.unam.mx*

** *Director del Instituto de Investigación y Desarrollo Educativo /UABC.*

Los procedimientos para evaluar la consecución de un conjunto de objetivos varían de acuerdo con los propósitos de los mismos. La empresa educativa requiere de configurar procedimientos sui generis para evaluar apropiadamente el alcance de sus metas.

Hasta hace poco, la mayoría de las evaluaciones escolares practicadas en México se limitaba a la elaboración de pruebas informales que los profesores hacían, las que comúnmente se utilizaban

sólo en el salón de clases para otorgar una calificación a sus estudiantes. Prácticamente no ha existido en las instituciones educativas una preocupación, menos una exigencia, por requerir que los profesores en servicio sean conocedores de la teoría y práctica de la medición educativa, dejando la evaluación de los alumnos a su buen entender.

Sin embargo, el papel de la evaluación ha cambiado drásticamente en los últimos años. De ser una práctica de poca importancia social, se ha convertido en una actividad de alto impacto, porque se empieza a tomar como evidencia, ya no sólo del nivel académico de los estudiantes, sino de la efectividad del maestro, la escuela y del mismo sistema educativo. Además, porque los resultados de la evaluación ahora se consideran como un elemento para orientar la toma de decisiones muy importantes, tales como la selección del ingreso de los alumnos, los estímulos al desempeño de los profesores y la asignación de presupuesto diferenciado a las escuelas y universidades.

Por esto debemos asegurar, ahora más que nunca, que la forma de medir o evaluar el rendimiento escolar de los estudiantes sea realizada con instrumentos de alta calidad que arrojen resultados válidos y confiables, que orienten la toma de decisiones correctas.

La tarea de evaluar implica la acción de valorar, determinar o estimar el valor de una cosa o de la resultante de una acción. En ello siempre hay un arbitraje, subyace un aspecto arbitrario, que no obedece a principios, leyes o propiedades intrínsecas a la naturaleza sino a la aceptación de un arbitrio dictado por la razón y la lógica de ideales que son socialmente aceptados por convicción de su conveniencia. Así, se tiene por convicción que la empresa educativa deberá generar las condiciones que permitan a los estudiantes la apropiación de un conjunto definido de conocimientos, habilidades, valores y actitudes.

Poder evaluar la empresa educativa, su eficacia (aptitud) y eficiencia (competencia), es un propósito del interés público de primera importancia. Esta preocupación universal, porque está en el interés de todas las naciones, implica una tarea ardua y compleja, que requiere de una investigación rigurosa y sistemática, para lograr los medios de evaluación que sean los más válidos y confiables. Este estudio reporta un trabajo de investigación sistemática dirigido a mejorar los procedimientos de evaluación del aprendizaje.

Un aspecto crucial de todo instrumento de evaluación es que evalúe lo que intenta evaluar. Resulta obvio que una prueba debe medir lo que pretende, sin embargo esto no es fácil de lograr, además es muy difícil saber si esto se ha conseguido o, mejor dicho, conocer hasta qué grado se ha alcanzado. Esto se conoce en la literatura clásica especializada de la psicometría como el problema de la validez de una prueba (Adkins, 1969; Cronbach, 1970; Magnusson, 1969; Nunnally, 1987).

Si se puede estimar la validez de un instrumento de evaluación, se puede tener en cuenta los alcances y limitaciones de nuestros procedimientos, lo que permite interpretar resultados y tomar decisiones ponderadas.

El problema más difícil para validar una prueba estriba en poder contar con otros medios que nos permitan evaluar la efectividad de la misma. Pongamos un ejemplo para ilustrar el problema. Si se desea apreciar qué tan buen estudiante llegará a resultar un aspirante a ingresar a la universidad, se puede hacer un examen de selección, el cual resultará tan efectivo como sea capaz de pronosticar a los buenos estudiantes; de aquí que el examen será válido si llega a diferenciar, con antelación y de manera atinada, a aquellos que logran obtener matas y buenas calificaciones en sus estudios profesionales. Por ello un procedimiento tradicional para validar una prueba es su valor predictivo, es decir, qué tan adecuada es su predicción en torno a la ejecución de las personas, en un área determinada de habilidades y conocimientos.

Entre más correlacionen los promedios escolares del alumno con el de la prueba, mejor será su validez criterial, especialmente su poder predictivo para seleccionar a los mejores estudiantes. De aquí que todos aquellos elementos que contribuyan a mejorar esta correlación, deberán ser

incorporados. De acuerdo con los resultados de nuestra investigación, el ofrecer en un examen de opción múltiple la alternativa para responder "no sé", permite mejorar la validez predictiva, además de otras ventajas que a continuación se exponen y que constituyen el propósito sustantivo de este trabajo.

En nuestro ejemplo de los aspirantes universitarios, se antoja como una prueba irrefutable la relación o, más apropiadamente dicho, la correlación de los puntajes de la prueba y las calificaciones obtenidas por los alumnos evaluados en el primer semestre o año escolar.

Si bien es cierto que la correlación es un buen indicador de la validez de la prueba, no debe dejar de llamar la atención de que esto no es necesariamente del todo apropiado. Bien podría ocurrir que un estudiante obtenga puntajes muy altos en la prueba de selección y, posteriormente, muy malas calificaciones en sus estudios universitarios del primer año y esto no sea porque la prueba opera inadecuadamente sino porque actúan otros factores. Por ejemplo, podría ocurrir el fallecimiento del padre de familia de un muy buen estudiante, desencadenando una crisis económica y emotiva que causen, posteriormente, un bajo desempeño. Por esto es difícil obtener indicadores de validación y se debe tener mucho cuidado al analizar los resultados.

En el campo de las pruebas escolares de selección que permiten diagnosticar y pronosticar la ejecución de un estudiante, el aspecto de su validez es un asunto primordial. Sobre todo porque hay muchas situaciones en las que la demanda sobrepasa con mucho la oferta educativa, aun habiendo oportunidad para todos, la demanda no es uniforme ni equilibrada, dado que hay ciertas carreras profesionales que están saturadas, como en el caso de derecho, contaduría y medicina. Por esto, aun teniendo capacidad de ingreso, se requiere una selección, para lo cual se debe contar con los procedimientos más adecuados, justos y equitativos. Dentro de estos procedimientos es necesario, aunque no suficiente, contar con instrumentos que permitan predecir, de la mejor manera posible, los que tendrán las mejores oportunidades de éxito académico. El incluir la opción "no sé" persigue este propósito.

La opción "no sé" es una aportación original de los autores de este trabajo, ya que no hay tradición sobre el uso de esta alternativa en los exámenes de opción múltiple; al menos en la exploración que hicimos sobre las bases de datos que incluyen las del *Educational Resources Information Center* el *Current Index Journals in Education* y el *Resources in Education* (ERIC, diciembre 1997), no encontramos alguna referencia al respecto; al igual que el Centro Nacional de Evaluación (CENEVAL), quienes nos reportaron que tampoco encontraron antecedentes.

La opción "no sé"

Al tratar de evaluar los conocimientos y habilidades que posee una persona siempre habrá limitaciones para hacerlo, pero se debe, en todos los casos, elegir la mejor alternativa. Una forma muy generalizada y estandarizada de evaluación son las pruebas de opción múltiple, por ser aplicables a gran escala, objetivas en su proceso de calificación, económicas, válidas, confiables y apropiadas para valorar respuestas de la más alta complejidad, como las de "evaluación", de acuerdo con la taxonomía de Bloom (1975). Este tipo de preguntas se pueden formular haciendo que todas las opciones de respuesta sean verdaderas, aunque siempre deberá haber una opción que es, sin lugar a duda, la mejor respuesta. Con este tipo de preguntas se tiene como consigna "evaluar" cuál de las opciones resulta la mejor. Considérese el siguiente ejemplo:

México tiene fronteras con: a) Estados Unidos b) Guatemala c) Estados Unidos y Guatemala
d) Estados Unidos, Belice y Guatemala e) No sé

En este ejemplo queda claro que todas las *respuestas son verdaderas*, sin embargo la opción "d" es la mejor por ser la más inclusiva y precisa.

Por otro lado, la opción "no sé" se agrega en los reactivos de opción múltiple como una invitación al estudiante para que se comporte con honestidad con respecto a lo que ignora y sabe.

El problema del factor de adivinación (guessing) en las pruebas de opción múltiple ha sido ampliamente analizado por diversos autores, dada su importancia en este tipo de instrumentos de evaluación (Brown, 1992; Budescu y Bar, 1993; Higbie, 1993; Talmir, 1991 ; Van der Ven y Gremmen, 1992; Wang, 1995; Zaremba y Schultz 1993); lo que hace palpable la importancia y vigencia de este asunto.

Hasta el momento, hemos encontrado 16 razones de por qué es conveniente agregar en los reactivos de opción múltiple la alternativa "no sé", mismas que a continuación se describen. Las primeras once aluden a una lógica educativa; las restantes se apoyan en razones psicométricas y se presentan resultados empíricos para sustentar las tesis que se plantean.

Razones educativas

El fin último de cualquier acción educativa, incluyendo la evaluación, es formar a los estudiantes: ya sea en su desarrollo intelectual, de habilidades, actitudes o valores. Por ello es importante considerar que en toda acción de evaluación debe privar la finalidad educativa. Promover la simulación en una prueba, o desestimularla, tiene consecuencias en la formación del estudiante, lo que se debe tener presente.

1. Un aspecto muy importante de por qué es apropiado invitar al alumno a que acepte y manifieste no saber cuando ignora la respuesta, es promover su sentido de responsabilidad para asumir sus limitaciones, además de que es ético alentar su honestidad al evitar incurrir en apariencias y falsedades.

2. El fin sustantivo de todo examen de conocimientos es poder estimar, con el mayor grado de precisión posible, el nivel de dominio que tiene la persona evaluada; por lo cual la apariencia de que se sabe, cuando en realidad se ignora, debe evitarse, y de no ser esto factible, al menos atenuarse en lo posible. Todos sabemos que obtener aciertos, tanto por responder adivinando como por copiar la respuesta del vecino en un examen, resulta un engaño que invalida los resultados de una prueba. Para evitar, o mejor dicho atenuar, esta situación, resulta apropiado invitar y ofrecer la opción para manifestar "no sé" cuando se ignora o no se recuerda la respuesta.

3. Uno de los problemas que más se enfatizan sobre los reactivos de opción múltiple, es que presentan una alta incidencia de respuestas azarosas que "contaminan" los resultados de la evaluación. Para atenuar este problema hay algunas estrategias que se pueden utilizar. Una de ellas consiste en restar un puntaje por cada error cometido, generalmente basado en la probabilidad del azar (0.25 en el caso de cuatro opciones), advirtiéndole al encuestado de que trate de evitar los errores para eludir dicha penalidad, es decir, persuadiéndole de que deje de contestar si no sabe. Esta estrategia la utilizan exámenes de gran escala y experiencia, como el *Scholastic Aptitud Test* (SAT) y el *General Management Admission Test* (GMAT), ambos elaborados por el *Educational Testing Service* (ETS), que es una institución de larga trayectoria y muy acreditada en la elaboración de pruebas.

Si la intención de la penalidad es que el alumno no conteste cuando no sabe, para evitar el factor "contaminante" de aciertos por azar, sorprende que en las opciones no se incluya de manera explícita y sistemática una alternativa para expresarlo, como "no sé", lo que haría más persuasivo, por estar presente, su uso.

4. En cualquier condición de evaluación, cuando no se tiene la habilidad y el conocimiento requeridos, lo más conveniente y deseable es que sea explícito, para evitar confundir o engañar al evaluador, por lo que es conveniente tener la opción "no sé".

5. La evaluación tradicionalmente ha centrado la atención en los aciertos, porque es lo único que se cuenta, dejando de lado el análisis y potencialidades del tipo de errores u omisiones que se generan. Por ejemplo, resulta obvio que es mejor reconocer cuándo se desconoce el camino y, por lo mismo, ocuparse en indagar la ruta, que quien toma una dirección incorrecta y tiene que corregir regresando lo andado. Es decir, saber qué se ignora es mejor que creer que se sabe o estar confundido. De aquí que responder "no sé" es muy diferente a cometer un error. Estas consideraciones pueden agregar un valioso recurso para afinar los procedimientos de evaluación.

6. En algunos exámenes de opción múltiple se le indica al estudiante que conteste todas las preguntas, incluso invitándolo a aventurarse cuando éstas se ignoran. El motivo de esto se basa en el supuesto de que no se puede evitar la respuesta azarosa y por lo tanto es mejor alentar a que todos tengan la misma oportunidad de obtener el beneficio de la suerte. Este argumento no es correcto porque agrega una ventaja diferencial a favor de los que menos saben, ya que son ellos los que tendrán mayor oportunidad de obtener aciertos adivinando. Si bien es cierto que no se puede evitar que se aventuren respuestas aleatorias en busca de aciertos, también es verdad que sí se pueden atenuar, primero, desalentándolas al establecer consecuencias adversas para quien incurre en esta práctica; segundo, promoviendo la respuesta "no sé", ofreciéndola siempre como una opción adicional.

7. Otro argumento que se ofrece a favor de que se invite a no dejar pregunta sin contestar, es que en la vida siempre hay que tomar riesgos y por lo mismo el castigar los errores inhibe esta iniciativa. Estamos de acuerdo en que es importante saber tomar riesgos, pero esto no quiere decir que hay que aventurarse frente a cualquier imponderable. Lo apropiado y formativo es enseñar a tomar riesgos cuando vale la pena hacerlo, saber distinguir cuando la probabilidad de éxito es alta o cuando es mejor abstenerse evitando errores que pueden ser muy costosos. Esto se puede favorecer al penalizar de manera gradual los que se cometen en una prueba. Por ejemplo, los tres primeros errores pudieran no recibir penalización en la calificación, en los subsecuentes tal vez penalizar la calificación desagregando 0.05 puntos por cada error de manera acumulativa, es decir, por el cuarto error se pierde 0.05 puntos, por el quinto 0.10, por el sexto 0.15, y así sucesivamente. Con este tipo de estrategia se le haría ver al estudiante que lo más conveniente es arriesgar, cuando más, a cometer un número pequeño de errores, siendo más apropiado contestar "no sé", con lo que se evita perder puntos.

8. Hay quienes consideran que al agregar la opción "no sé" se amplía el número de elecciones que tiene que hacer el estudiante, lo que puede resultar inconveniente. Pero de hecho en toda prueba de opción múltiple se puede dejar de contestar, lo que implica que siempre hay una posibilidad más de las que se ofrecen como alternativas de respuesta, aunque esto no es explícito. Cuando se penalizan los errores, como algunas pruebas lo hacen, lo que promueven es la respuesta de omisión para cuando se ignora la contestación. En cualquier condición, con o sin castigo, los estudiantes evaluados utilizan "no sé", de manera indirecta, omitiendo la respuesta, o directa, si se ofrece la opción, por esto resulta más conveniente hacerla explícita.

9. Otra ventaja de la opción "no sé" es que puede contribuir a identificar las respuestas de azar que emitió un estudiante al contestar la prueba, con lo cual se aumenta la validez de los resultados. Esto se puede ilustrar muy bien poniendo situaciones hipotéticas extremas. Considérese que un alumno, primer caso, obtiene: 25% de aciertos, 75% de respuestas incorrectas y nunca opta (0%) por contestar "no sé"; en cambio hay otro, segundo caso, que obtiene: 25% de aciertos, 0% de respuestas incorrectas y opta el 75% de las ocasiones por contestar "no sé". O considérese un tercer caso que obtiene: 75% de respuestas correctas y el 25% restante son errores; a diferencia de un cuarto caso que obtiene: 75% de respuestas correctas y el 25% restante se pronuncia por "no sé". Queda claro que tanto el primero como el tercer caso están contaminados por el azar, y que el segundo y cuarto no lo están, o al menos proporcionalmente así es.

10. Incluir la opción "no sé" puede constituir un buen factor de ponderación para calibrar la evaluación, se puede promover su uso sin tener que recurrir al castigo por los errores cometidos. Esto es posible haciendo saber al examinado que a igualdad de aciertos se estimará como mejor

ejecución quien tenga el mayor número de omisiones (respuestas "no sé"), de manera tal que si, por ejemplo, en un examen de admisión tres alumnos igualan su calificación obteniendo 48 aciertos, pero el primero tiene nueve omisiones, el segundo cinco y el tercero sólo una, tendrán preferencia de ingreso los que tengan el mayor número de "no sé".

Este procedimiento se torna en una manera de premiar a quien asume responsablemente cuando ignora la respuesta y se evita añadir una penalidad por incurrir en errores, además, cómo ya se señaló en la novena razón, quien cuenta con más "no sé" y tiene el mismo número de aciertos, resulta razonable considerar que su calificación está menos "contaminada" por respuestas acertadas al azar.

11. El azar también se puede conocer con las pruebas que contienen reactivos estructurados por áreas de conocimiento y niveles de dificultad, al analizar cuándo el estudiante contesta las preguntas difíciles y falta en las fáciles o cuándo manifiesta no saber en las fáciles y contesta las difíciles; lo que además permite valorar la coherencia interna de las respuestas del alumno.

Hasta aquí hemos expuesto una serie de razones formalmente pensadas, pero hay otros argumentos que son más contundentes para recomendar el uso de la opción "no sé", porque éstos se desprenden de ventajas que se sustentan en la evidencia empírica, es decir, de los resultados de investigación sistemática, observando y analizando las respuestas emitidas de "no sé" en pruebas de opción múltiple.

Razones psicométricas

Los resultados que se reportan en este trabajo se basan en diversas investigaciones realizadas por los autores, fundamentalmente en los análisis obtenidos de las pruebas de validez que aplicamos al Examen de Habilidades y Conocimientos Básicos (EHXCOBA) (Backhoff y Tirado; 1992, 1993; Tirado, Backhoff, Larrazolo, y Rosas; 1997), prueba que se aplicó durante el ciclo 94-95, en la Universidad Autónoma de Baja California a aproximadamente 9000 aspirantes.

Con el objeto de conocer la validez predictiva del EHXCOBA se correlacionan las calificaciones obtenidas en esta prueba con el promedio de las del primer semestre o año. De igual manera se correlacionan los promedios obtenidos en el bachillerato con la calificación obtenida en el EHXCOBA. El número de estudiantes que integraron la muestra en este estudio fue de 1 293 casos.

12. El uso de la opción "no sé" correlaciona positivamente con el nivel de dificultad de la prueba, es decir, a mayor dificultad más recurrente es el uso de esta opción, como era de esperarse. En la tabla 1 se muestran 20 reactivos (numerados del 1 al 20) del área de ciencias sociales del EHXCOBA, donde la correlación entre el porcentaje de aciertos y el porcentaje del uso de la opción "no sé" es de -0.78.

Aunque no todas las secciones de este examen presentan esta correlación tan alta, todas tienen correlaciones estadísticamente significativas (tabla 2).

Tabla 1

Porcentaje de aciertos, errores y "no sé" en una muestra de reactivos del EXHCOBA

Reactivos	Aciertos	Errores	No sé
1	20	46	34
2	52	44	3
3	35	57	8
4	38	50	12
5	41	45	14
6	67	31	2
7	42	42	16
8	39	48	13
9	36	48	16
10	27	51	22
11	56	32	12
12	59	26	15
13	62	33	6
14	52	39	9
15	50	43	7
16	49	48	3
17	80	18	2
18	80	18	2
19	89	8	3
20	68	31	2

Tabla 2

Correlaciones de la dificultad de los reactivos con el uso de la opción "no sé" en algunas secciones del EXHCOBA

Secciones	Preguntas	Correlación
Habilidades verbales	30	0.60
Habilidades cuantitativas	30	0.60
Español	15	0.37
Matemáticas	15	0.68
Ciencias naturales	20	0.42
Ciencias sociales	20	0.78

13. El uso que se hace de la opción "no sé" depende de la dificultad y de la conformación del reactivo, presentando un comportamiento diferenciado coherente y con variaciones muy notorias. Por ejemplo, en la aplicación de una prueba para evaluar conocimientos de biología con una muestra compuesta por 962 alumnos, se dieron variaciones de magnitud muy significativas: en un reactivo optaron por aceptar no conocer la respuesta hasta un 71.1% de los evaluados, mientras que en otra pregunta sólo el 0.9% marcaron "no sé" (Tirado y López, 1994). Este comportamiento tan claramente diferenciado hace patente que la opción "no sé" constituye una alternativa que es utilizada de manera selectiva, y que puede mejorar sustancialmente la capacidad de evaluación de una prueba, porque si se hubiera invitado a los examinados a responder aun sin saber, como

ocurre en algunas evaluaciones, aproximadamente una cuarta parte del 71.1 % de los que indicaron no saber hubieran, por suerte, acertado ya que de las cuatro opciones una es correcta (probabilidad = 0.25); introduciendo así una importante suma de aciertos tergiversados, afectando con ello la validez y la confiabilidad de la prueba.

14. También es importante destacar que en la evidencia experimental acumulada, como en el ejemplo antes descrito, se observa que hay patrones claros y consistentes de respuestas asociados con la opción "no sé", lo que permite definir características en los reactivos; con lo cual se constituyen elementos valiosos para la investigación de los procesos de evaluación.

15. Como ya se apuntó en la octava razón, se ha argumentado que la inclusión de la opción "no sé" incrementa el número de alternativas, haciendo más difícil el reactivo, lo que en realidad no ocurre ya que la posibilidad de dejar de responder está de hecho siempre presente; pero también se debe señalar que si se coloca la opción "no sé" sistemáticamente como la última y en todos los reactivos, entonces se hace muy estable y por lo mismo no agrega para el examinado diversidad ni dificultad, como hemos podido apreciar al encuestar a los examinados. Además Crehan (1993) reporta que preguntas con menos opciones de respuesta (tres en lugar de cuatro) resultan ser más difíciles.

16. Como último argumento, deseamos exponer uno de los criterios más sólidos para sustentar la conveniencia de usar sistemáticamente la opción "no sé", el cual se basa en evidencia empírica, mostrando la posibilidad de mejorar la validez predictiva de una prueba, lo que es de suma importancia en la elaboración de test psicométricos y de evaluación, importancia que ya hemos sustentado en otro trabajo (Tirado, Backhoff, Larrazolo y Rosas; 1997).

Al correlacionar la calificación del Examen de Habilidades y Conocimientos Básicos con el promedio de bachillerato, así como con el de las calificaciones obtenidas en el primer semestre de estudios universitarios y el primer año, estos valores fueron de 0.34, 0.37 y 0.39, respectivamente.

Tomando como premisa que quienes en un mayor número de ocasiones respondieron "no sé" es más probable que sus aciertos estén menos "contaminados" por respuestas al azar, se decidió ajustar la calificación obtenida en el EXHCOBA restando una fracción de punto (0.25) por cada error cometido, dejando sin afectar la calificación en el caso de haber evitado cometer un error (quienes marcaron: "no sé"). El valor promedio del EXHCOBA fue de 14.96 aciertos, de un máximo de 190, el cual bajó a 98.82, una vez que se aplicó las restas por cada error cometido, lo que corresponde a un ajuste equivalente al 14.03% sobre la calificación promedio original. Con la calificación ajustada, las correlaciones mejoraron con las calificaciones promedio del bachillerato, del primer semestre y del primer año, obteniendo los valores de 0.35, 0.39, y 0.41, respectivamente, que son estadísticamente significativos ($p < 0.001$).

Si bien es cierto que la magnitud del alza en la correlación fue de una y dos centésimas de punto, también es cierto que estos valores representan una mejora relativa sobre los originales que van de 2.86% a 5.13%, lo que no es despreciable, porque se trata de un procedimiento que permite ajustar la calificación de una prueba incrementando su validez predictiva, lo que es de suma importancia y muy difícil de lograr.

Conclusiones

Probablemente pudiera haber algunas otras razones de por qué puede resultar conveniente introducir de manera sistemática la opción "no sé"; también algunas o muchas de las dieciséis expuestas en este trabajo, pudieran ser apreciadas por algunos lectores como de poca importancia. No obstante esto, es muy conveniente observar que no vemos ninguna razón sustantiva para no incluir la opción "no sé", por lo que parece plenamente recomendable hacerlo.

Mejorar los valores de validez predictiva de una prueba es, a todas luces, deseable por lo que parece muy conveniente utilizar un procedimiento como el descrito en este trabajo. Pero aquí no se

agotan las posibilidades para mejorar las pruebas de opción múltiple introduciendo la alternativa "no sé", hay un sinnúmero de posibles análisis utilizando esta variable como factor de ponderación, por lo que existe un campo potencial inexplorado.

Cada vez es más patente y difundida la necesidad de evaluar, y para ello se deben afinar todos los procedimientos que permitan hacerlo de la manera más equitativa, justa, válida y confiable. Las pruebas de opción múltiple son sólo una posibilidad dentro de un amplio espectro de procedimientos de evaluación, pero su importancia es crucial porque permite su aplicación a gran escala. En tanto mejoremos estos instrumentos, podremos seguir avanzando y, por lo mismo, tomar decisiones más apropiadas.

Referencias bibliográficas

Adkins, W. D. (1969). *Elaboración de tests*, México: Trillas.

Backhoff, E. y Tirado, F. (1992). "Desarrollo del Examen de Habilidades y Conocimientos Básicos", en Revista de la *Educación Superior*, vol. XXI, núm. 83, pp. 95-118, julio-septiembre, México: Asociación Nacional de Universidades e Instituciones de Educación Superior.

Backhoff, E. y Tirado, F. (1993). "Habilidades y conocimientos básicos del estudiante universitario: hacia los estándares nacionales", en Revista de la *Educación Superior*, vol. XXII, núm. 88, pp. 45-65, octubre-diciembre, México: ANUIES.

Bloom, B. S. (1975). *Taxonomía de los objetivos de la educación*, Buenos Aires: El Ateneo.

Brown, J. R. (1992). "Trial by fire one perspective for guessing", en *Psychology in the Schools*; vol. 29, núm. 1, pp. 71-77.

Budescu, D. y Bar H.M. (1993). "To guess or not to guess: a decision theoretic view of formula scoring", en *Journal of Educational Measurement*, vol. 30, núm. 4, pp. 277-291.

Crehan, K. D. (1993). "Use of an inclusive option and the optimal number of options for multiple-choice items", en *Educational and Psychological Measurement*, vol. 53, núm. 1, pp. 241-247.

Cronbach, L. J. (1970). *Essentials of psychological testing*, Nueva York: Harper International.

Higbie, J. (1993). "Multiple choice exams: how well do we know the result?", en *Physics Teacher*, vol. 31, núm. 7, pp. 426-427.

Magnusson, D. (1969). *Teoría de los tests*, México: Trillas.

Nunnally, C. J. (1987). *Teoría psicométrica*, México: Trillas.

Talmir, P. (1991). "Multiple choice items: how to gain the most out of them", en *Biochemical Education*; vol.19, núm. 4, pp.188-191.

Tirado F. y López Trujillo A. (1994). "Evaluación de la enseñanza de la biología en México, hacia una transformación en la calidad de la educación", en Revista de la *Educación Superior*, vol. XXIII, núm. 89, pp. 165-189, enero-marzo, México: ANUIES.

Tirado, Backhoff, Larrazolo, y Rosas (1997). "Validez predictiva del Examen de Habilidades y Conocimientos Básicos (EHXCOBA)", en *Revista Mexicana de Investigación Educativa*, vol. ii, núm. 3, enero-junio, México: COMIE /UNAM.

Van der Ven, A. H. y Gremmen, F. M. (1992). "The knowledge or random guessing model for matching tests", en *Applied Psychological Measurement*, vol.1 6, núm. 2, pp. 177-194.

Wang, J. (1995). "Critical values of guessing on true-false and multiple-choice tests". Paper presented at the *Annual Meeting of the American Educational Research Association*, San Francisco.

Zaremba, S. B. y Schultz, M. (1993). "An analysis of traditional classroom assessment techniques and a discussion of alternative methods of assessment". *Teaching of Psychology: Ideas and Innovations*. Proceedings of the Annual Conference on Undergraduate Teaching of Psychology, 7th, Ellenville, Nueva York.

Recepción del artículo: 7 de octubre de 1998

Aceptado: 15 de enero de 1999