

Evaluación de un sistema de evaluación del aprendizaje en psiquiatría

Gerardo Campo-Cabal¹

Resumen

Introducción: Mediante el análisis del sistema de evaluación del aprendizaje utilizado en un curso de psiquiatría ofrecido a estudiantes de medicina, este artículo revisa elementos básicos de todo proceso de evaluación del aprendizaje. *Métodos:* Análisis de los métodos de evaluación utilizados y las calificaciones obtenidas por los estudiantes de cada uno de los cuatro grupos en que se divide el grupo total de alumnos. *Resultados:* Los métodos de evaluación seleccionados son apropiados para evaluar los objetivos educacionales; el contenido es definido mediante una matriz de especificaciones; el coeficiente de correlación alto entre las calificaciones acumuladas en los semestres académicos anteriores y las alcanzadas en el curso, y entre las calificaciones (considerando tanto la totalidad como solo una parte del examen) son evidencia de la validez de los resultados. La curva de calificaciones desviada a la derecha demuestra que la mayor parte de los alumnos adquiere los conocimientos esperados. *Conclusión:* El sistema de evaluación utilizado en el curso de psicopatología es justo, válido y confiable, en especial en lo referente a los métodos objetivos utilizados, pero debe mejorar la *evaluación conceptual* o, preferiblemente, eliminarse como parte constitutiva del sistema de evaluación.

Palabras clave: Evaluación educacional, educación médica, psiquiatría, instituciones de enseñanza superior

Title: Assessment of an Evaluation System for Psychiatry Learning

Abstract

Through the analysis of a teaching evaluation system for a Psychiatry course aimed at Medicine students, the author reviews the basic elements taken into account in a teaching assessment process. *Methods:* Analysis was carried out of the assessment methods used as well as of the grades obtained by the students from four groups into which they were divided. *Results:* The selected assessment methods are appropriate to evaluate educational objectives; the contents are selected by means of a specification matrix; there is a high correlation coefficient between the grades obtained in previous academic periods and the ones obtained in the course, thus demonstrating the validity of the results (both considering the whole exam or just a part of it). Most of the students are on the right side of the grading curve, which means that the majority of them acquire the knowledge expected. *Conclusions:*

¹ Psiquiatra y Magíster en Educación Médica, Profesor Titular, Jefe de Departamento de Psiquiatría de la Universidad del Valle, Cali, Colombia.

The assessment system used in the Psychopathology course is fair, valid and reliable, specifically concerning the objective methods used, but the *conceptual evaluation* should be improved or, preferably, eliminated as a constituent part of the evaluation system.

Key words: Educational measurement; education, medical; psychiatry; higher education institutions

Introducción

La asignatura Psicopatología es un curso teórico ofrecido a los estudiantes del quinto semestre de Medicina de la Universidad del Valle, con metodología de clase magistral en cada uno de los cuatro grupos en que se divide el casi centenar de alumnos. La evaluación del aprendizaje comprende una evaluación parcial (40%), una evaluación final (50%) y una evaluación conceptual (10%); esta última, producto de la observación del desempeño del estudiante en cada una de las clases programadas. Una semana después de cada evaluación se realiza un nuevo examen, la evaluación opcional del parcial y evaluación opcional del final, y el estudiante opta por la calificación más alta en cada caso (1-4).

El año inmediatamente anterior a la realización del presente estudio, la coordinación del curso de Psicopatología le fue asignada a un docente vinculado recientemente al Departamento de Psiquiatría. Ante la concentración de calificaciones aprobatorias en los estudiantes asignados al grupo del coordinador y

las no aprobatorias en los restantes tres grupos, los alumnos afectados presentaron el correspondiente reclamo. Al revisar lo ocurrido, se encontró que el examen aplicado consistía en preguntas abiertas, sin que se hubiese elaborado una guía de los elementos y conceptos que deberían incluir las eventuales respuestas correctas; de lo anterior se desprende que el marcaje de la prueba quedaba expuesta al sesgo del evaluador.

Como parte de los correctivos instaurados, al año siguiente se reasignó la coordinación a un profesor con mayor experiencia y formación docente. Para sorpresa de los miembros de la unidad académica, tras aplicarse la primera prueba escrita, *examen parcial*, las críticas a la evaluación provinieron no de los estudiantes, sino de dos de los profesores del curso, uno de los cuales era el profesor relevado de la coordinación. Dada la situación planteada, se consideró necesario proceder a evaluar cuán válidas y pertinentes eran las quejas expresadas por los colegas profesores y limitar al ámbito pedagógico la resolución del conflicto académico planteado. El presente artículo pretende brindar los argumentos para dirimir tal conflicto, con base en el análisis del sistema de evaluación y las calificaciones obtenidas por los estudiantes en el curso de Psicopatología.

Al revisar la literatura sobre evaluación del aprendizaje es posible encontrar que los expertos en

educación sugieren que para lograr un sistema de evaluación del aprendizaje justo es necesario considerar las siguientes preguntas: ¿para qué evaluar?, ¿cuándo evaluar?, ¿quién debe evaluar?, ¿cuáles deben ser las características de la evaluación?, ¿qué debe ser evaluado?, ¿cómo evaluar?, y adicionalmente definir las características de los exámenes, su marcaje y el nivel de exigencia de la competencia; además, comunicar de manera adecuada el resultado del proceso de evaluación (5,6).

Richardson propone como propósitos de la evaluación obtener información de lo apropiado del contenido de un determinado curso, de la efectividad del proceso de enseñanza-aprendizaje, y establecer una medida del nivel de desempeño del estudiante. Adicionalmente, la evaluación puede servir para: seleccionar entre los aspirantes a un programa académico, medir la calidad del sistema educativo, organizar subgrupos clasificándolos con base en el nivel de los alumnos, diagnosticar una situación para poder emprender correctivos, considerar el alcance de objetivos intermedios y pronunciarnos sobre el éxito o fracaso de todo el proceso, determinar si se puede otorgar o no la respectiva certificación, y, muy importante, la función predictiva sobre el desempeño futuro.

La evaluación aplicada antes del proceso de enseñanza aprendizaje permite valorar el nivel con el que el estudiante arriba a la situación educativa; posteriormente, puede

aplicarse durante todo el proceso o en algún momento intermedio, y finalmente, al término del proceso, para medir el logro de los objetivos propuestos (7).

Existe una tendencia en las últimas décadas de propiciar la participación en el proceso de evaluación de nuevos actores, como los pares del sujeto evaluado, el personal administrativo, y en las profesiones de la salud, la participación del paciente directamente involucrado en las actuaciones del estudiante; pero en general podemos afirmar que la responsabilidad del proceso evaluativo continúa siendo fundamentalmente del profesor.

Definir qué va a ser evaluado puede ser guiado por el diseño del currículo, esto es, si el currículo está construido por objetivos educacionales o por competencias; y adicionalmente, por el momento en el cual se realiza la evaluación. Así, durante el proceso se puede intentar evaluar el avance del alumno, y al final, verificar la adquisición del aprendizaje necesario para un adecuado desempeño en las situaciones reales. En otras palabras, por objetivos educacionales se pueden evaluar los conocimientos, destrezas y actitudes; y por competencias, constatar el saber, hacer y el ser del individuo.

Existen diversos instrumentos de evaluación que se clasifican según estén orientados hacia los procesos o hacia los productos, lo que permite aproximarse a la escogencia

de cómo realizar la evaluación con base en el propósito definido para la evaluación. Para la evaluación durante el proceso se mencionan los portafolios, las entrevistas, las observaciones documentadas, los registros de aprendizajes diarios, la autoevaluación, los informes de entrevistas sobre proyectos, productos y muestras de alumnos, y la expresión en voz alta del pensamiento del alumno; mientras que para evaluar los productos encontramos instrumentos como las pruebas de ensayo extensa o limitada, los proyectos, las carpetas de trabajos, los muestreos, las investigaciones, los productos de expresión artística y los tests estandarizados del tipo “selección múltiple”.

Los instrumentos de evaluación deben seleccionarse y los resultados de su aplicación se deben considerar con base en los criterios de validez, confiabilidad, practicidad y utilidad. El grado de precisión con que se mide lo que se desea medir corresponde a la validez. Más que la validez del instrumento, lo que interesa es la validez de los resultados y el uso que se hará de estos.

Existen diferentes clases de validez: la de contenido, cuando una muestra significativa del universo del contenido cubierto por el curso o la unidad didáctica está representado; la validez predictiva, que permite prever el desempeño posterior del alumno en aspectos que corresponden al área evaluada; la validez de construcción, cuando existe coherencia entre el instrumento de

evaluación y el marco teórico que sustenta el proyecto pedagógico; la validez de convergencia, cuando se establece la relación que existe entre un programa de evaluación o un instrumento aplicado, y otros programas o instrumentos de validez ya conocida; la validez manifiesta, que se refiere al modo en que los instrumentos aparecen frente al público externo, quien los encuentra razonables; la validez de significado, que corresponde a la relación que se establece desde la perspectiva de los estudiantes entre el programa y los instrumentos de evaluación, por un lado, y los procesos de enseñanza y aprendizaje, por el otro; y la validez de retroacción, cuando la evaluación tiene un efecto normativo sobre los contenidos de la enseñanza, esto es, la evaluación establece lo que se debe enseñar, por ejemplo, lo ocurrido con los resultados obtenidos por los estudiantes de medicina de Colombia en el examen de Estado Ecaes y las adecuaciones curriculares realizadas por varias universidades del país con el propósito de mejorar el rendimiento de sus estudiantes.

El grado de exactitud con que se mide un determinado rasgo corresponde a la confiabilidad; debe ser estable y objetiva, esto quiere decir que, independientemente de quien utiliza un programa o un instrumento de evaluación, se obtengan los mismos resultados. La practicidad se refiere a la viabilidad de la construcción, administración y análisis de resultados; y la utilidad, al potencial orientador

sobre las acciones por seguir con base en los resultados obtenidos.

Finalmente, señalar que al aplicar una evaluación pueden presentarse varias situaciones: en la *situación menos afortunada*, todos los estudiantes fallan en adquirir los conocimientos, y al graficar estos datos, la curva se mantiene desviada a la izquierda; en la *situación ideal*, el proceso de enseñanza aprendizaje permite que todos los alumnos adquieran los conocimientos esperados y la curva de calificaciones se desvía a la derecha; pero lo que *habitualmente* se presenta es una distribución normal que discrimina el diferente nivel de aprendizaje, donde aquellos que pese al proceso de enseñanza no adquieren el aprendizaje se mantienen a la izquierda, un grueso mayoritario del grupo evidencia haber adquirido el conocimiento, al obtener calificaciones aprobatorias, y un pequeño grupo, el de los estudiantes distinguidos, logran calificaciones sobresalientes que corresponden al extremo derecho de la curva.

Ahora bien, cuando se cometen errores en la construcción de los métodos de evaluación puede ocurrir que estudiantes sin el aprendizaje apropiado sean sometidos a una prueba excesivamente fácil, lo que da como resultado que las calificaciones sean inmerecidamente altas, o viceversa, estudiantes competentes sometidos a un examen mal construido o extremadamente difícil, derivan en calificaciones injustamente bajas (7,8).

Métodos

Ante las críticas aparecidas después de aplicar el examen parcial, se decide que lo más apropiado es considerar el *sistema de evaluación del aprendizaje* del curso de Psicopatología como un objeto educativo, susceptible de ser evaluado por medio de un estudio observacional (8). Evaluación posible si se procede al análisis de los métodos de evaluación del aprendizaje utilizados (examen escrito y evaluación conceptual) y de las calificaciones obtenidas por los estudiantes en cada uno de los cuatro grupos.

En referencia al primer punto, en los métodos de evaluación se revisan las consideraciones tenidas en cuenta para la construcción de los exámenes escritos, por ejemplo: qué se pretende evaluar con la prueba, la elaboración de una tabla de especificaciones que garantice un muestreo representativo de los contenidos del curso, la adecuada formulación de las preguntas y que el nivel de dificultad de estas corresponda a estudiantes del tercer año de Medicina (9,10). En referencia a la tradicionalmente llamada *evaluación conceptual*, se pregunta a cada uno de los profesores participantes en el curso si disponen de un instrumento estructurado que especifique las variables por considerar y que sirvan de referente para emitir la calificación de cada estudiante en los diferentes grupos (11,12).

Posteriormente, se realiza un estudio y análisis estadístico de las

calificaciones obtenidas por los estudiantes como un todo y como participantes en cada uno de los cuatro grupos: A, B, C y D. Finalmente, estas calificaciones son correlacionadas con las calificaciones alcanzadas por los estudiantes en los cuatro semestres previamente cursados, para intentar establecer la validez del sistema de evaluación empleado y explicar las eventuales diferencias en el desempeño entre los estudiantes (13).

Resultados

La evaluación del aprendizaje se realiza en varios momentos, a mitad y final del semestre en lo que se refiere a la evaluación escrita de conocimientos, y en un proceso continuo en lo referente a la evaluación conceptual. Se aplican 77 exámenes parciales y 76 exámenes finales, y el mismo número de pruebas en las sendas evaluaciones opcionales. Un estudiante del grupo D canceló la asignatura, y por ello no presentó las evaluaciones finales, ni es incluido en el análisis estadístico.

Los métodos de evaluación fueron escogidos tomando en consideración los objetivos educacionales definidos para el curso, que pueden resumirse en la adquisición de conocimientos, más que en el logro de habilidades o en la demostración de determinadas actitudes. El 90% de la calificación de cada estudiante se deriva del resultado

en las pruebas (parcial y final), de allí la importancia de garantizar la adecuada construcción de los exámenes escritos. Los exámenes escritos incluyen preguntas de selección múltiple, selección múltiple modificada, respuestas cortas y de correlación.

Las preguntas demandan de los estudiantes recordar hechos y procedimientos, por ejemplo, evocar datos relevantes de la historia de la psiquiatría, citar los elementos característicos de la conducta anormal, mencionar las secciones constituyentes de la historia clínica psiquiátrica o del examen mental, etc.; y la comprensión de conceptos y principios, entre ellos comprender y discriminar entre conceptos como sensación y percepción, poder discriminar entre un trastorno perceptivo con o sin la presencia de un estímulo real, comprender la relación entre vivencias a edad temprana y la ulterior aparición de síntomas frente a la influencia de la sociedad en la génesis de la enfermedad, etc.

Una matriz o tabla de especificaciones para cada uno de los exámenes permite garantizar la inclusión de todos los contenidos cubiertos en el curso, y la ponderación considera el tiempo del curso dedicado a cada uno de los tópicos. La tabla 1 ilustra la matriz de especificaciones de un examen de 30 preguntas aplicado como examen parcial.

Tabla 1. Matriz de especificaciones. Primer parcial, curso de Psicopatología, Departamento de Psiquiatría, Universidad del Valle

Contenido	Nº de preguntas	Ponderación	Datos	Procedimientos	Conceptos	Principios
Historia	5	16,6%	5	-	-	-
Conducta anormal	3	10%	1	-	2	-
Modelos	2	6,6%	1	-	1	-
Técnica de entrevista	5	16,6%	2	2	1	-
Historia clínica	5	16,6%	3	2	-	-
Semiología	10	33,3%	3	3	3	1

Nota: Ponderación determinada por el tiempo del curso dedicado a cada contenido. La mitad de las preguntas requieren del estudiante la evocación de datos y hechos concretos; aproximadamente una cuarta parte explora procedimientos y una proporción igual, la comprensión de conceptos, lo que guarda coherencia con los objetivos educacionales para la primera parte de este curso.

La primera crítica por parte de los dos profesores giraba en torno a considerar irrelevante que el 16,6% de las preguntas exploraran conocimientos de la historia de la psiquiatría, poco pertinente en su concepto para un médico general; y agregaban que ello explicaría el pobre resultado alcanzado por los estudiantes de sus respectivos grupos.

Al tomar en consideración esta observación, se procedió a marcar nuevamente el examen; se encontró que la calificación obtenida por cada estudiante no difiere de manera significativa al considerar o no las preguntas de la historia de la psiquiatría, o sea que el *error estándar de medida* es pequeño. Adicionalmente, se encuentra una correlación positiva alta al comparar la calificación obtenida de las dos maneras, y el hecho de que el número de estudiantes (15 estudiantes) con calificaciones no aprobatorias

(menor de 3,0 sobre 5,0) no se modifica de una u otra manera. Por otra parte, solo dos de los 77 estudiantes mejoran la calificación aprobatoria inicial (entre 3,0 y 3,9) a aprobatoria sobresaliente (4,0-5,0) (figura 1).

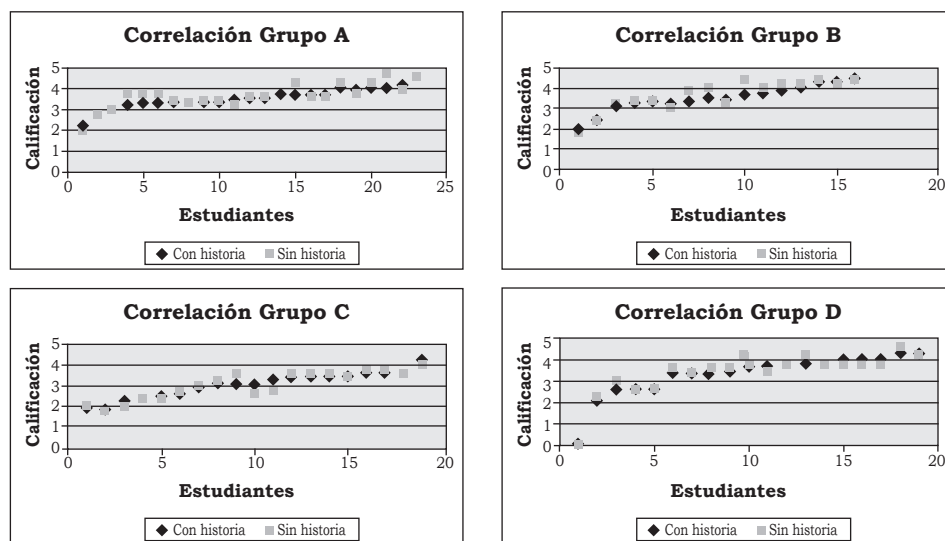
En cada uno de los cuatro grupos se confirma que considerar o no las preguntas sobre historia de la psiquiatría no afecta de manera estadísticamente significativa las calificaciones obtenidas por cada uno de los estudiantes. Adicionalmente, vale la pena anotar que no existe diferencia estadísticamente significativa al comparar el promedio aritmético de las dos maneras de marcar, en cada uno de los grupos: grupo A: 3,5 frente a 3,6; grupo B: 3,5 frente a 3,6; grupo C: 2,9 frente a 3,0; grupo D: 3,4 frente a 3,5 (no se incluye un estudiante que no presentó el examen) (figura 2).

Figura 1. Calificaciones con y sin preguntas de la historia de la psiquiatría



Nota: El coeficiente de correlación entre las calificaciones es alto: 0,92.

Figura 2. Correlación entre calificaciones con y sin historia en cada grupo



Nota: Un estudiante del grupo C y otro del grupo D logran la calificación aprobatoria mínima, pero sin que exista diferencia estadísticamente significativa entre las calificaciones originales y las calificaciones finales. Dos estudiantes del grupo C que habían logrado calificaciones aprobatorias pasan a tener calificaciones no aprobatorias. Coeficientes de correlación: A: 0,87; B: 0,88; C: 0,92; D: 0,97.

Dos estrategias han sido utilizadas para definir el nivel de exigencia de una prueba: un estándar relativo (*norma-referenciado*), donde el nivel de desempeño depende del

desempeño de todos los estudiantes; o un estándar absoluto (*criterio-referenciado*), donde independiente del desempeño de los evaluados, con base en el análisis del contenido de

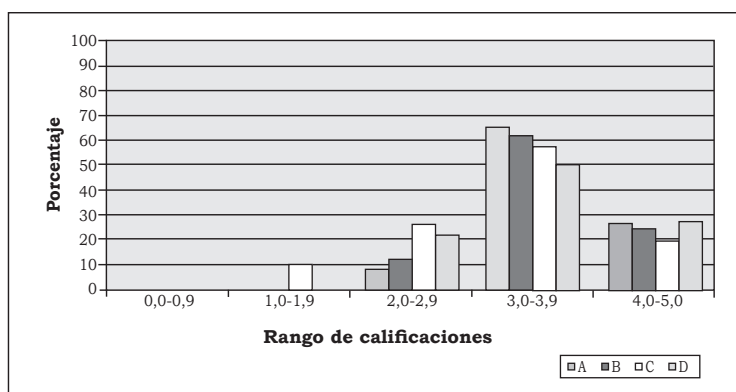
la prueba se escoge arbitrariamente un porcentaje de respuestas correctas necesarias para pasar (14); este último criterio es el utilizado en la Universidad del Valle, donde es necesario obtener una calificación mayor o igual a 3,0 sobre un máximo posible de 5,0 para aprobar.

Ante la segunda crítica expresada por los profesores a cargo de los grupos C y D, respecto a un examen “mal construido” y con una dificultad extremadamente alta, se controvierte fácilmente al considerar las calificaciones obtenidas por la totalidad de los estudiantes que presentan una curva de distribución normal, ligeramente desviada a la derecha. El 20,7% obtiene calificación no aprobatoria, incluyendo el estudiante que no presentó el examen; el 58,4%, calificación aprobatoria, entre 3,0 y 4,0; y el 20,7%, calificación aprobatoria-sobresaliente,

correspondiente a mayor o igual a 4,0; para un total de 79,2% de estudiantes que aprobaron el examen.

Con la intención de evaluar el tercer señalamiento de un posible sesgo a favor de los estudiantes del grupo A, correspondiente al coordinador del curso, se comparan las calificaciones obtenidas en los cuatro grupos. Las calificaciones más altas se encuentran en el grupo D, si bien no difiere de manera significativa con los grupos A y B. La proporción de estudiantes con calificaciones no aprobatorias en los grupos C y D, en efecto, es mayor que en los otros dos grupos, pero las calificaciones no aprobatorias de los grupos A y B no difieren entre sí de manera estadísticamente significativa, por lo cual la explicación de las diferencias observadas con el grupo D habría que buscarlas en otras variables intervinientes en el proceso de enseñanza aprendizaje (figura 3).

Figura 3. Calificaciones por grupos (A, B, C, D) en el primer parcial



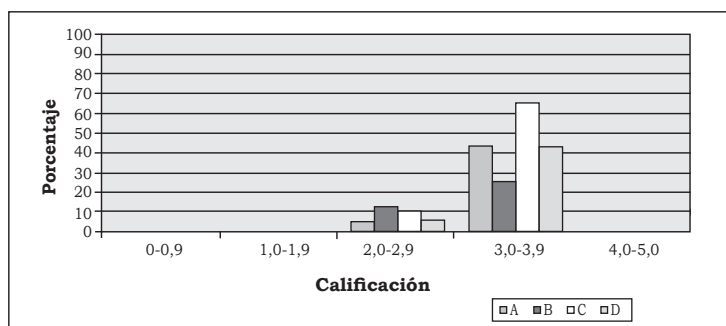
Nota: Los grupos A y B presentan resultados semejantes. Solo el grupo C presenta calificaciones no aprobatorias entre 1,0 y 1,9: 10,52%. El estudiante que no presentó el examen es incluido en el grupo D. Los grupos A, B y D presentan calificaciones aprobatorias-sobresalientes entre 4,0-5,0: 26,08%, 25% y 27,77%, respectivamente, frente a solo el 5,6% en el grupo C.

Una semana después es aplicado el opcional del primer parcial al 76,7% de los estudiantes que voluntariamente optan por esta posibilidad de mejorar su calificación. Considerando que entre la aplicación de uno y otro examen no se produce ningún contacto entre los estudiantes y sus profesores, las calificaciones obtenidas significativamente más altas en algunos alumnos podrían ser explicadas como el producto de: la presentación del primer examen como una experiencia de aprendizaje, la retroalimentación brindada por la calificación obtenida, eventuales actividades de autoaprendizaje. La

calificación asignada finalmente a cada estudiante corresponde a la más alta de las dos.

Los alumnos con calificaciones aprobatorias en la evaluación parcial pasan de 79,2% a 88,8% del total; el promedio aritmético del grupo mejora al pasar de 3,5 a 3,8 sobre 5,0 ($p < 0,05$); y que en los grupos A, B y D más de la mitad de los estudiantes obtengan calificaciones aprobatorias y sobresalientes (entre 4,0 y 5,0) debería ser interpretado como una prueba de que el propósito de lograr los objetivos educacionales propuestos se cumplió (figura 4).

Figura 4. Calificación definitiva del primer parcial



Nota: Calificación promedio del grupo total: 3,76. Calificación aprobatoria (3,0-5,0): grupo A, 95,6%; grupo B, 87,5%; grupo C, 90%; grupo D: 94,7%. Calificación aprobatoria-sobresaliente (4,0-5,0): grupo A, 52,2%; grupo B, 62,5%; grupo C, 25,0%; grupo D: 52,6%.

Con base en los resultados presentados hasta aquí se puede responder con hechos a las críticas expresadas; pero, con la intención de evaluar todo el sistema de evaluación del curso de Psicopatología, se realiza un análisis semejante a los otros exámenes suministrados (final

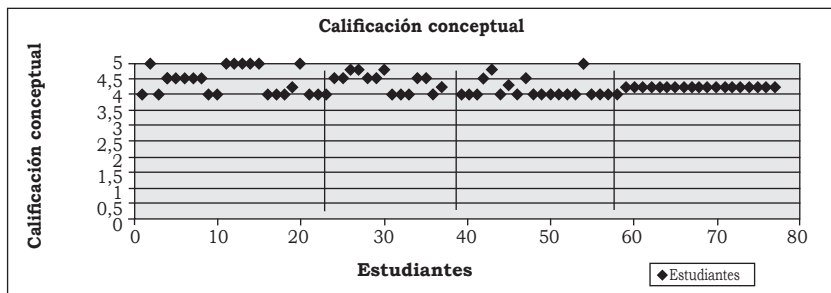
y opcional del final), con resultados comparables, pero que por la extensión de este informe son omitidos.

No fue posible unificar criterios para la evaluación conceptual, por las conceptualizaciones tan divergentes encontradas en los cuatro profesores participantes, que van

desde expectativas tan limitadas como verificar la asistencia a las clases programadas, pasando por la pasiva y “respetuosa” atención exhibida durante las clases magistrales, hasta la ambiciosa y poco probable de que los estudiantes pudiesen demostrar activamente los conocimientos adquiridos. De manera que la evaluación conceptual emitida en

los cuatro grupos (A: 23 estudiantes; B: 16 estudiantes, C: 19 estudiantes, y D: 19 estudiantes) es subjetiva y corresponde a la particular interpretación de cada profesor, sin que alguno de ellos utilizara un instructivo como referencia. Las calificaciones producto de la evaluación conceptual se muestran en la figura 5.

Figura 5. Evaluación conceptual



Nota: Todos los estudiantes obtienen calificación entre 4,0 y 5,0. En los grupos A y C se asignan calificaciones de 5,0. A todos los estudiantes del grupo D se les asigna igual calificación: 4,2.

En referencia a las calificaciones definitivas, el grupo A mantiene el buen resultado inicial en las calificaciones definitivas, evidenciado por el alto promedio aritmético de las evaluaciones (3,96), si bien no presenta diferencias estadísticamente significantes con el promedio logrado por los estudiantes del grupo D (3,95). El grupo B, pese a mostrar un muy buen logro de sus estudiantes, es el que menos progresión evidencia entre las evaluaciones parciales y las calificaciones definitivas, lo que escapa a poder ser explicado con los resultados

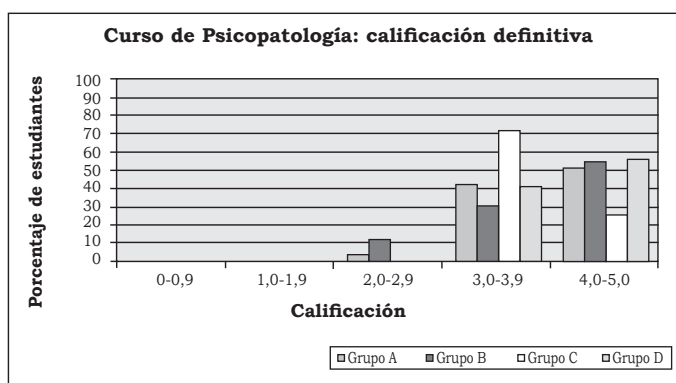
disponibles. En el grupo C, pese a presentar la más baja proporción de estudiantes con calificaciones aprobatorias-sobresalientes, todos sus miembros obtienen una calificación aprobatoria. Estos dos grupos, el B y el C, presentan los promedios de calificaciones más bajos (3,78 y 3,75). Todos los estudiantes de los grupos C y D aprueban, mientras que en los grupos A y B reprueban uno y dos estudiantes, respectivamente. En el grupo C, pese a que ningún estudiante reprueba, es en el que se presenta una menor proporción de calificaciones aprobatorias-sobre-

salientes. Pero, además, es evidente que en los grupos A, B y D existe una proporción de estudiantes con calificaciones sobresalientes que no difieren de forma estadísticamente significativa (26,1%, 25,0% y 27,8%, respectivamente) entre ellos.

Ahora bien, la crítica puede surgir de la alta proporción de estudiantes con calificaciones no aprobatorias en los grupos C y D

(36,8% y 22,2%), en comparación con el grupo A (8,6%); pero, como ya se mencionó, sin que este último difiera estadísticamente del grupo B (12,5%); y sin dejar de anotar que un estudiante del grupo D no presentó la evaluación escrita, por lo cual se le asigna una calificación igual a 0.0 y contribuye a inflar la proporción de estudiantes que reprueban el examen (figura 6).

Figura 6. Calificación definitiva en el curso de psicopatología

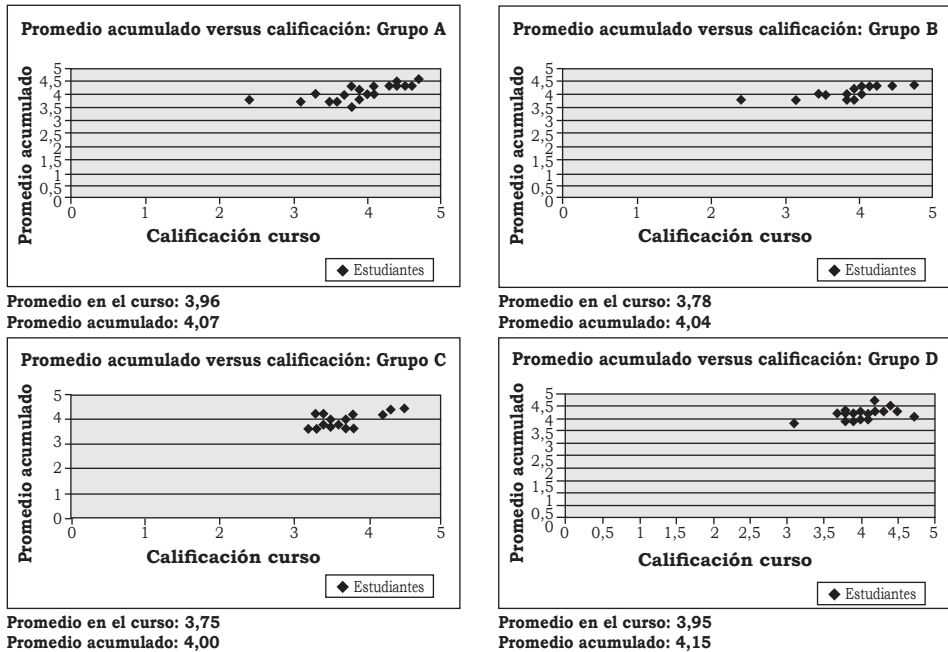


Nota: Calificaciones aprobatorias sobresalientes en su orden: grupos D (57,9%), B (56,3%), A (52,2%) y C (26,3%).

Análisis

Finalmente, en la figura 7 se presenta una correlación entre las calificaciones obtenidas en cada uno de los grupos del curso y el promedio acumulado de cada estudiante en los semestres precedentes; los cuales, para el total del grupo, es de 4,07 sobre 5,0, frente a un promedio aritmético en el curso de Psicopatología de 3,87 sobre 5,0.

La revisión del sistema de evaluación del aprendizaje en el curso de Psicopatología nace como respuesta a las críticas expresadas por dos de los profesores participantes en el curso, quienes, tras aplicarse el primer examen escrito (evaluación parcial), consideran que este presentaba tres fallas: irrelevancia al incluir dentro del examen los con-

Figura 7. Correlación: calificación del curso y calificación acumulada en Medicina

tenidos de historia de la psiquiatría; en segundo lugar, el examen habría quedado mal construido y con un grado de dificultad demasiado alto; finalmente, un sesgo a favor de los estudiantes del grupo A, que correspondía al profesor responsable de la evaluación del aprendizaje. Todo lo cual habría afectado negativamente el rendimiento de sus respectivos estudiantes.

Si bien la primera crítica debería ser objeto de una profunda reflexión alrededor de lo que debe ser la formación integral de los médicos, este no es el propósito del presente trabajo, por lo tanto mantenemos el foco en evaluar el juicio formulado por los colegas profesores, y se demuestra fehacientemente que el resultado

no se modifica de forma estadísticamente significativa al eliminar las preguntas referentes a la historia de la psiquiatría, con una correlación de confiabilidad positiva y alta entre los dos puntajes (los coeficientes de correlación más altos precisamente en los grupos C y D), con un error de medida muy estrecho; o sea que la mayoría de estudiantes obtienen calificaciones de una y otra forma que no difieren de manera estadísticamente significativa al comparar los promedios aritméticos de las calificaciones de cada grupo, y ello es evidente al comparar los promedios totales (promedio aritmético, 3,35 y 3,46). Adicionalmente, la elaboración de una tabla de especificaciones garantiza que cada uno de los temas

cubiertos en el curso es tenido en cuenta.

Ante la segunda observación, el coordinador evidencia que en la construcción del examen las preguntas incluidas corresponden a una adecuada muestra de los contenidos cubiertos en la primera parte del curso y que fueron diseñadas de forma que exigían de los educandos no solo la mera evocación de datos y procedimientos, sino, también, la comprensión de conceptos y principios. De igual manera, el señalamiento del nivel desproporcionadamente alto de dificultad carece de fundamento, evidenciado en que aproximadamente el 80% del total de estudiantes logran calificaciones aprobatorias, y con calificaciones aprobatorias-sobresalientes por al menos uno de cada cuatro alumnos en los grupos A, B y D.

Los estudiantes del grupo C presentan un comportamiento estadísticamente diferente, con solo el 5,6% con calificación aprobatoria-sobresaliente, y el 36,8% con calificaciones no aprobatorias, que incluyen un 10,5% con calificaciones entre 1,0 y menores a 2,0, en un rendimiento del alumnado muy pobre. Este, al no poder ser explicado por la dificultad del examen aplicado, indicaría la necesidad de investigar otras variables intervinientes en el proceso pedagógico para que no se haya producido el aprendizaje esperado.

La última crítica en referencia a un posible sesgo en favor de los miembros del grupo A parecería confirmarse para un observador

desprevenido, por cuanto este grupo tiene el más alto porcentaje de estudiantes que aprueban el primer parcial (91,3%), pero al revisar los resultados cuidadosamente se evidencia que este grupo no se diferencia estadísticamente del grupo B (87,5%). Adicionalmente, la mayor concentración de estudiantes con calificaciones aprobatorias-sobresalientes está en el grupo D (27,8%); curiosamente, un grupo a cargo de uno de los dos profesores que presentaron quejas.

Por todo lo enunciado, es posible afirmar que los tres señalamientos críticos carecen de fundamento académico, y que el pobre resultado alcanzado por los alumnos del grupo C no es explicable por un alto nivel de dificultad de la prueba escrita, o por su mala construcción, o que esta tuviese un sesgo negativo en contra de ellos; más bien, habría que buscar la explicación en alguna otra de las variables que intervienen en el proceso de enseñanza aprendizaje.

Como se mencionó antes, el examen opcional del primer parcial es realizado una semana después sin que medie un nuevo encuentro entre los alumnos y los profesores, por lo que es llamativa la mejoría notable en los resultados definitivos en el primer parcial logrados por los estudiantes del grupo C, quienes, en una proporción superior al 90%, logran calificaciones aprobatorias; si bien en todos los grupos se observa algo semejante, reflejado por ejemplo en que la proporción de estudiantes

que obtienen calificaciones aprobatoria-sobresaliente (entre 4,0 y 5,0) supera el 50% en los grupos A, B y D; y en el grupo C, pese a continuar con los más pobres resultados, esta proporción se cuadriplica (5,6% a 26,3%). Otro hallazgo interesante es la persistencia de la proporción de estudiantes del grupo B que re-prueban el examen parcial (12,5%), situación que se mantiene hasta el final del curso, pese a los buenos resultados alcanzados por los otros estudiantes de este grupo.

Los resultados correspondientes a la evaluación final y evaluación opcional del examen final no son presentados por la extensión que requeriría el presente informe, pero basta saber que las calificaciones se correlacionan positivamente con las obtenidas en el primer parcial, y que los resultados del examen opcional final les permite a muchos de los estudiantes mejorar su calificación definitiva.

La evaluación conceptual, como se esperaba al no considerar un instrumento estructurado que especifique las características por evaluar de parte del observador, no permite una evaluación objetiva y presenta un claro sesgo a favor de los estudiantes, donde todos ellos obtienen un 4,0 como mínima calificación (sesgo de tendencia central). Vale la pena resaltar cómo el profesor del grupo D asigna a todos los estudiantes idéntica calificación, 4,2, lo que seguramente premia inmerecidamente a algunos alumnos, y castiga injustamente a

otros que probablemente sí cumplieron con las acciones esperadas de ellos, como leer el material con anterioridad, asistir a las sesiones programadas, participación activa durante las sesiones, etc.

Al comparar los resultados obtenidos en la evaluación parcial y la calificación definitiva obtenida por los estudiantes en el curso de Psicopatología se observa que la mejor evolución corresponde a los estudiantes del grupo D, entre quienes no se presentan calificaciones no aprobatorias y se concentra la mayor proporción de alumnos con calificaciones aprobatorias-sobresalientes.

Con base en las calificaciones obtenidas por los estudiantes en los primeros cuatro semestres (promedio acumulado: 4,07) y el alcanzado por los estudiantes en el curso de Psicopatología, se encuentra que el desempeño en el curso fue ligeramente inferior (promedio del curso: 3,87). En el grupo D se encuentra el promedio acumulado más alto (promedio: 4,15) y en el grupo C, el promedio más bajo (promedio: 4,0), pero sin que existan diferencias estadísticamente significantes entre los cuatro grupos que pudieran explicar el diferente nivel de desempeño presentado por los estudiantes del grupo C en el primer parcial o el más bajo promedio aritmético de las calificaciones definitivas de los grupos B y C.

Las anteriores consideraciones no explican los resultados de los tres alumnos que reprobaban el curso, ni

el desempeño apenas aceptable de los estudiantes del grupo C, por lo cual es necesario el diseño de un estudio con variables diferentes a las consideradas en la presente investigación, que permitan responder a estos interrogantes. Por ejemplo, el profesor del grupo A, interesado por el bajo rendimiento del estudiante que reprobó en su grupo, pudo establecer que este presentaba la reagudización de un trastorno mental severo, y que en oposición a las recomendaciones del médico tratante y del profesor mismo, el estudiante decide no retirarse del curso.

En síntesis, los exámenes escritos empleados en el curso de Psicopatología miden de manera válida y confiable el grado de conocimientos alcanzado por cada uno de los estudiantes. La mitad de los alumnos (36/77) evidencia sobresalientes conocimientos en los hechos, procedimientos, conceptos y principios tratados en el curso; y la otra mitad ha demostrado su adquisición, si bien estos conocimientos deberán ser reforzados en las posteriores rotaciones clínicas por el Departamento de Psiquiatría.

Es imperativo mejorar la evaluación conceptual, la cual, como queda demostrado, tiene fallas innegables, si bien es necesario aclarar que ningún estudiante obtuvo en el curso una calificación aprobatoria con base en esta.

En conclusión, el presente trabajo es un ejemplo de cómo las

discrepancias académicas deben ser abocadas y resueltas con base en argumentos académicos. En la situación específica de la evaluación del aprendizaje del curso de Psicopatología, y con base en los criterios propuestos por expertos en el tema, se pudo corroborar que el sistema de evaluación utilizado cumple con la tarea de evaluar los conocimientos a los estudiantes del curso de manera válida y confiable; que los exámenes escritos aplicados están bien contruidos y permiten discriminar el nivel de aprendizaje, pero que la evaluación conceptual requiere su inmediata revisión y mejoramiento, o, mejor aún, su eliminación del sistema de evaluación.

Referencias

1. Campo-Cabal G. Curso de psicopatología. En: Descripción cursos de pregrado. Cali: Universidad del Valle; 2008.
2. Campo-Cabal G, Castrillón E, Bersh S, et al. Curso de Psicopatología. Material Impreso, Departamento de Psiquiatría. Cali: Universidad del Valle; 2007.
3. Campo-Cabal G, Álvarez J, Morales AM. La entrevista médica con un enfoque terapéutico. *Rev Col Psiquiatr.* 2006;35:547-69
4. Universidad del Valle. Consejo Superior, Acuerdo 009: Reglamento estudiantil. Capítulo IV: Del proceso de evaluación. Cali: Universidad del Valle; 1997.
5. Harden RM. Assess students: an overview. *Medical Teacher.* 1979;1:65-70.
6. Broadfoot PM. Education, assessment and society. Harry Torrance. University of Sussex (Series Editor) of Assessing Assessment. Buckingham, Philadelphia: Open University Press; 1996.

7. Richardson R. Diploma in medical education. Centre for Medical Education, Dundee DD2 1LR. Assessment and Evaluation. Scotland, UK. 1997.
8. Campo-Cabal G. Evaluation of the assessment system of the fifth year medical students at Universidad del Valle with view to improvement. A dissertation submitted to the Centre for Medical Education, in partial fulfillment of the requirement for a Master Degree in Medical Education. University of Dundee, Scotland 1997-1998.
9. Harden RM. ASME Medical Education Booklet N°10, constructing multiple choice questions of the multiple true/false type. Dundee: Association for the study of Medical Education; 1979.
10. Campo G, Melo E, Isaza C, et al. Examen escrito de conocimientos en Ciencias Básicas-curriculum tradicional y currículo innovador. Presentación de resultados. Cali: Universidad del Valle; 1998.
11. Whitfield CF, Xie SX. Correlation of problem-based learning facilitators' scores with student performance on written exams. *Advances Health Sci Educ.* 2002;7:41-51.
12. Newble DI. ASME Medical Education Booklet N° 25. Assessing clinical competence at the undergraduate level. *Medical Education.* 1992;26:504.
13. Brown G, Bull J, Pendlebury M. Chapter 15: Reliability, validity and examining, En: Brown G. *Assessing student learning in higher education.* London: Ed. Routledge; 1997.
14. Gronlund NE, Linn RL *Measurement and evaluation.* Alaska: Peacock Publishers; 1990.

Conflictos de interés: El autor manifiesta que no tiene conflictos de interés en este artículo.

Recibido para evaluación: 20 de abril de 2012

Aceptado para publicación: 18 de mayo de 2012

Correspondencia
Gerardo Campo-Cabal
Jefe del Departamento de Psiquiatría
Universidad del Valle
Calle 5ª No.36-00
Cali, Colombia
gercampo@univalle.edu.co