

Universidad de Ciencias Médicas de Holguín
Hospital Pediátrico "Octavio de la Concepción de la Pedraja"

Calidad de exámenes certificativos de Pediatría. Municipio Holguín.
Curso 2012 – 2013.

Autora: Dra. Yanexy Pérez González.

Tutor: Dr. Pedro Augusto Díaz Rojas.

Memoria escrita para optar por el título de Master en Educación Médica

2014

Dedicatoria

A mi madre por lo que representa

A mis hijos por su cariño

A mi esposo por su amor

A mi familia por lo que es

Agradecimientos

A mi tutor Dr. Pedro Augusto Días Rojas por haberme preparado en el terreno de la educación e investigación médicas, por su acertada guía en la realización de este trabajo y su constante estímulo.

Al colectivo de profesores de la Maestría en Educación Médica por el caudal de conocimientos aportados.

A todas aquellas personas que me acompañaron para que un principio, tuviera un final.

ÍNDICE

Resumen

I.	Introducción.....	1
II.	Marco Teórico.....	6
III.	Objetivos.....	26
IV.	Diseño Metodológico.....	27
V.	Análisis y Discusión de los Resultados.....	36
VI.	Conclusiones.....	60
VII.	Recomendaciones.....	61
VIII.	Referencias Bibliográficas.....	62

Anexos

RESUMEN

La presente investigación se realizó con el objetivo de caracterizar la calidad de los exámenes certificativos de Pediatría aplicados a estudiantes del cuarto año de la carrera de medicina del municipio Holguín, en el curso 2012 – 2013. Se emplearon métodos teóricos, empíricos y procedimientos estadísticos como expresión de la triangulación metodológica. Se describieron las características de los exámenes según su estructura, temas, áreas de aprendizaje exploradas y tipos de preguntas e incisos utilizados, apoyados en los resultados de una entrevista a especialistas de la asignatura. Se identificó la correspondencia entre los contenidos de la unidad curricular y su exploración en los exámenes y se valoró, en conjunto con los resultados del grupo focal de estudiantes y profesores, el comportamiento del índice de dificultad, el índice y coeficiente de discriminación del punto de correlación biserial y el coeficiente alfa de Cronbach como indicadores de calidad de los instrumentos. Los resultados evidenciaron que ambos exámenes fueron mixtos con predominio de las preguntas test objetivo, se incluyeron todas las temáticas, la clínica fue el área de aprendizaje más explorada y se encontró total correspondencia entre los contenidos de la unidad curricular y su exploración en los instrumentos, corroborado por el coeficiente de correlación de Spearman. El examen II fue el de mayor grado de dificultad, con la sepsis e inmunidad como temática más difícil, según el coeficiente de discriminación el de mayor poder predictivo y solo en este existió correspondencia entre los incisos esperados y reales. Ambos exámenes tienen una confiabilidad aceptable.

Descriptores: evaluación, exámenes, Pediatría

INTRODUCCIÓN

La evaluación es tan antigua como la vida misma, en todo momento y circunstancias ha habido algún tipo de evaluación: pequeña o grande, espontánea o programada, científica o por tanteo. De la bondad de la misma, depende en buena medida la seguridad y el bienestar de la persona evaluada y de la que evalúa.

La evaluación se hace imprescindible en todos los ámbitos de la vida, especialmente en el área de la educación, donde controlar, evaluar los resultados, tomar decisiones, son operaciones constantes en el proceso enseñanza aprendizaje. Por todo esto la problemática de la evaluación es un problema de la educación ya que docente y alumno se interrelacionan de modo tal que, lo que aquel haga, repercute inevitablemente en el discípulo.¹

En la evaluación del aprendizaje los exámenes escritos, son una de las varias herramientas que forman parte de un amplio abanico de opciones evaluativas que tienen carácter de sistema, y tienen como función general la regulación del proceso docente educativo.²

Son mundialmente los de mayor uso, los más económicos y los de mayor flexibilidad técnica por lo que son escogidos en nuestras universidades médicas para las evaluaciones teóricas que certifican el final de una rotación o asignatura. Ellos complementan el examen práctico y en conjunto con la evaluación formativa y/o educativa determinan el éxito o fracaso del estudiante.²

Diversos autores como Chadwick, Rotger, Villarroel, Camperos, reflejaron una fuerte crítica a los métodos tradicionales de evaluación, y en especial a los exámenes de papel y lápiz, por sus formalismos y limitaciones.²

Si a lo antes expuesto le sumamos las numerosas propuestas de nuevos diseños de entornos evaluativos constructivistas con fuertes apoyos en las nuevas tecnologías de la información; más el auge que han tomado otras modalidades como la evaluación del desempeño, la evaluación auténtica, la de portafolios, entre otras que han ido introduciéndose progresivamente en el ámbito educativo.

¿Los exámenes escritos, el más antiguo y universal de los instrumentos evaluativos, no tienen ya espacio dentro de las nuevas concepciones educativas, por su incapacidad para responder a los objetivos y propósitos previstos?²

Se ha podido identificar un grupo de dificultades que permiten comprender por qué dichos instrumentos son frecuentemente considerados “estereotipados y tradicionales”. Lo que demuestra sin dudas que en materia de examen, no todo está hecho.

Algunas de las características frecuentes de los exámenes escritos, así lo demuestran:²

- Carencia de un diseño estéticamente pertinente (agradable, motivante).
- Cuestionarios, muchas veces inclinados a serie de ítems simplistas, que exigen predominantemente respuestas estrechas, fijas e irreflexivas. En los que no hay espacio para la creatividad.
- Están dirigidos, más a comprobar los conocimientos acumulados, que a propiciar el desarrollo y el descubrimiento de procedimientos y funciones mentales de los estudiantes puestas en acción durante su realización.
- No están convenientemente orientados hacia la solución de las problemáticas principales existentes en el aprendizaje, y muchas veces son ajenos a tareas reales en que se desempeñará el futuro profesional.
- No son suficientemente contextualizados e integradores.
- No son coherentes con el resto de las tareas docentes ni con el sistema evaluativo del que forman parte.
- Se limitan o acogen casi exclusivamente a un criterio de estandarización. No atienden la diversidad psicológica, los distintos niveles de desarrollo alcanzado por los alumnos, ni brindan opciones para las diversas necesidades de aprendizaje.
- No establecen verdaderos retos al aprendizaje (carencia de rigor científico).
- No son reconocidos suficientemente como medio a través del cual también se instruye, se educa y se moviliza el interés por estudiar más y mejor. Por tanto, no son habitualmente desarrolladores.

Todas estas características hacen del examen escrito un instrumento afuncional a partir del cual es difícil impulsar, valorar y mejorar los aprendizajes de los alumnos. Pero, es obvio que los problemas no están en el instrumento, sino más bien en las concepciones tradicionales desde las cuales estos son elaborados, aplicados y calificados. El secreto del éxito está más bien, en el perfeccionamiento continuo de las técnicas e instrumentos y en su complementación racional.

En los últimos años existe la preocupación por indagar sobre la eficiencia del proceso de evaluación, así como también su influencia en el proceso enseñanza aprendizaje del alumno.

Numerosos son los trabajos realizados en el ámbito nacional e internacional, en todos los niveles de educación, desde la básica hasta la universitaria, de materias teóricas como prácticas, que coinciden en que la evaluación y dentro de ella la confección de los instrumentos requiere de una atención especial. Destacan la profunda necesidad de asesoramiento académico en cuanto a los aspectos involucrados en el proceso de evaluación de los aprendizajes.

Según Liliana Ortiz en su artículo “La nueva concepción de la evaluación de aprendizajes en la educación de las Ciencias de la Salud”, en Chile existe la premisa entre los estudiantes de que la evaluación habitualmente es injusta e indican que los instrumentos de evaluación utilizados por los docentes no miden el conocimiento que deben alcanzar de las materias y que las calificaciones no reflejan lo que saben. Opinan que la mayoría de los docentes de las carreras profesionalizantes como Medicina, poseen escasa o nula formación pedagógica por lo tanto carecen de competencias evaluativas.³

La premisa, entre los académicos, es que los estudiantes no han aprendido suficiente en los cursos previos. Señalan que los instrumentos de evaluación, que habitualmente se diseñan se refieren a un número muy reducido de competencias cognoscitivas, muchas veces reducida a la memorización comprensiva y dejan a un lado un conjunto importante de procesos y competencias involucrados en el aprendizaje que debieran ser objeto de evaluación.³

Calidad de los exámenes certificativos de Pediatría.

Múltiples han sido los estudios que en estos últimos años se han realizado acerca de la calidad de los exámenes escritos en nuestras universidades y la médica no queda atrás. Con frecuencia encontramos exámenes con marcado carácter reproductivo y memorístico, poco creativos, ajenos a situaciones reales a las que se puede enfrentar el futuro profesional, todo lo cual hace cada vez más empírica y subjetiva la evaluación.

En la actualidad existen un grupo de profesores e investigadores cubanos que han dedicado parte de su trabajo a este tema entre los que se destacan Salas Perea, Hatím, Díaz Rojas entre otros.^{4-8.}

Un ejemplo de ello es el estudio de calidad del examen diagnóstico de la asignatura Morfofisiología Humana I realizado por Carrazana y Salas Perea en el que se reconoció por parte de estudiantes y profesores la necesidad de estos exámenes diagnósticos para corregir dificultades en la evaluación, mejorar el trabajo metodológico y reajustar la autopreparación.⁹

Reconocidos trabajos en esta rama se realizaron en la República Bolivariana de Venezuela por maestrantes cubanos de Educación Médica Superior, y que tienen continuidad en nuestro país y especialmente en nuestra provincia; quienes demuestran que aún falta calidad en su confección lo cual traduce deficiencias en la planeación de estos instrumentos.

La autora es del criterio que el proceso de planeación de la evaluación se hace necesario e imprescindible para minimizar los errores en la confección de los exámenes y con ello garantizar una mayor calidad de los mismos y un camino a la excelencia en el proceso docente educativo.

En la asignatura de Pediatría, como en el resto de las asignaturas de la carrera de medicina se realizan evaluaciones durante el proceso enseñanza aprendizaje que son formativas y certificativas. Las evaluaciones frecuentes y parciales son las formativas mientras que el examen final es el considerado como certificativo para la asignatura.

Calidad de los exámenes certificativos de Pediatría.

Esta evaluación final de la competencia clínica consta de una parte práctica y otra teórica. En el examen práctico se comprueba el nivel adquirido por el estudiante en las habilidades clínicas, mediante los diferentes modos de actuación médica.

En el teórico comprobamos el nivel de conocimiento alcanzado y la capacidad para enfrentar el proceso para la solución de problemas clínicos, sociomédicos e

higiénico-epidemiológicos en los pacientes en edades pediátricas y este lo realizamos de manera escrita.

Hace dos cursos académicos se realizó por primera vez, por la autora de este trabajo, una caracterización del examen final escrito de pediatría demostrándose falta de calidad en su confección pues no se exploraban los objetivos en su totalidad, y el nivel medio de dificultad para el examen quedó por debajo de los valores establecidos por diferentes autores como aceptables, con un predominio de las preguntas e incisos fáciles.

Estos resultados motivaron a la autora comenzar a rectificar el proceso de planeación de estos instrumentos dentro de la cátedra de pediatría pero a pesar de los esfuerzos realizados los estudiantes aún expresan insatisfacción con la evaluación escrita de la asignatura, alrededor de un 12 % piden revisión de exámenes y los profesores plantean el bajo grado de discriminación para delimitar entre los alumnos bien y mal preparados. Todas estas insatisfacciones hacen que surja la siguiente **interrogante científica**:

¿Qué calidad presentan los exámenes certificativos de pediatría aplicados en el cuarto año de la carrera de Medicina en los semestres del curso 2012-2013 en el municipio Holguín?

MARCO TEÓRICO

La enseñanza constituye el proceso de organización y dirección de la actividad cognoscitiva e incluye, por tanto, la actividad del profesor que es enseñar y la del educando que es aprender, y es por ello que se denomina proceso de enseñanza-aprendizaje, proceso docente-educativo o simplemente proceso docente.¹⁰

El proceso docente educativo.

Olga Landaluce refiere en su libro Pedagogía que cuando el proceso formativo se ejecuta con carácter de sistema se denomina proceso docente educativo y que la práctica histórico social ha demostrado que en este proceso es donde se produce la formación de las nuevas generaciones de acuerdo con las aspiraciones de la sociedad.¹¹

El proceso docente se subordina al encargo social y responde a él, por lo que el mismo tiene que desplegarse con un enfoque sistémico y dinámico. Para llevar a cabo un correcto trabajo educativo, debe tenerse siempre presente la interrelación dialéctica entre la educación, la enseñanza-aprendizaje y la instrucción.¹²

El proceso docente se concreta en la instrucción, que expresa el resultado de la interacción profesor-educando en cuanto a la asimilación de los sistemas de conocimientos y habilidades, la capacidad de aplicarlos en forma creadora, así como la formación y desarrollo de la concepción científica del mundo. Modela la personalidad del que aprende y le posibilita el desarrollo de nuevos modos de actuación que le permiten cumplir exitosamente sus funciones sociales.¹⁰

En el proceso de enseñanza-aprendizaje existen dos componentes: los personales y los no personales. En los primeros se incluyen el profesor y el alumno. En los componentes no personales o categorías didácticas encontramos el objetivo, el contenido, el método, el medio, la forma y la evaluación de la enseñanza.¹³

La autora coincide con varios investigadores del tema que son los objetivos la categoría rectora de este proceso, porque jerarquiza el resto de los componentes. Pero es la evaluación quien comprueba el grado de cumplimiento de estos objetivos

y nos permite conocer la utilización acertada del resto de los componentes del proceso para su confirmación o rectificación.

Evaluación.

Evaluar es una reflexión que analiza las causas y los factores que motivaron un desempeño, un rendimiento o una participación acertada o no, excelente o deficiente. El propósito no es identificar a los que sí tuvieron éxito o a quienes perdieron o fracasaron, sino de orientar o reorientar el trabajo de unos y otros; dicho análisis ha de hacerse individual y grupal a fin de dinamizar estrategias.

Evaluar no es una medición de conocimientos adquiridos u objetivos alcanzados, ya que no tiene sentido rechazar a quienes no lo logren en la proporción requerida como mínima. Pues, es necesario superar las barreras que afectan el ritmo personal del rendimiento de un educando, su autoestima y motivación.²

La evaluación del aprendizaje es el mecanismo regulador del proceso enseñanza aprendizaje, nos posibilita su retroalimentación, dirección y perfeccionamiento. Determina los modos de actuación que los estudiantes incorporan en su proceso formativo y acredita y certifica el cumplimiento de los objetivos terminales del futuro profesional y especialista.^{4, 5.}

El concepto evaluación para algunos aparece en el siglo XIX con el proceso de industrialización que se produjo en Estados Unidos En este marco surge el moderno discurso científico en el campo de la educación, que va a incorporar términos tales como tecnología de la educación, diseño curricular, objetivos de aprendizaje o evaluación educativa.¹⁴

Para otros la concepción aparece con los mismos comienzos de la sociedad la cual siempre ha buscado dar juicios de valor a las acciones y actitudes de los estudiantes y a través de los años muchos han sido los autores que se han detenido en el estudio de la evaluación y han emitido diversos conceptos y criterios sobre esta categoría por lo que la evaluación como disciplina ha sufrido profundas

transformaciones conceptuales y funcionales a lo largo de la historia y especialmente en el siglo XX y XXI.¹⁴

Pero quien tradicionalmente desde 1942 se considera como el padre de la evaluación educativa es Ralph Tyler por ser el primero en dar una visión metódica de la misma, pues define la evaluación como “el proceso que permite determinar en qué grado han sido alcanzados los objetivos educativos propuestos”. Plantea la necesidad de una evaluación científica que sirva para perfeccionar la calidad de la educación. La evaluación como tal desde esta perspectiva ya no es una simple medición por que supone un juicio de valor sobre la información recogida.¹⁵

Cronbach propone como alternativa una evaluación que se oriente fundamentalmente a buscar información y comunicarla a quienes han de tomar decisiones sobre la enseñanza, la define como “la recogida y uso de información para tomar decisiones sobre un programa educativo”.¹⁶

Pedro Lafouecade, argentino, doctor en Pedagogía y consultivo de la UNESCO expone: “La evaluación es una etapa del proceso educacional que tiene que comprobar de modo sistemático, en qué medida se han logrado los objetivos que se han especificado con antelación”.¹⁷

En 1996 García Ramos expone que la evaluación, en sentido general, es un conjunto de procesos sistemáticos de recogida, análisis e interpretación de información válida y fiable, que en comparación con una referencia o criterio nos permite llegar a una decisión que puede favorecer la mejora del objeto evaluado.¹⁸

Salas Perea en el 1998, define la evaluación como la actividad que consiste en el análisis e interpretación de los resultados de diferentes instrumentos de medición sobre la actuación de un educando o candidato ante una actividad y/o tarea, en comparación con un patrón o modelo estandarizado, y que se efectúa con la finalidad de tomar una buena decisión.⁴

Para Cerda Gutiérrez la evaluación es la acción de juzgar o inferir juicios sobre cierta información recogida directa o indirectamente de la realidad evaluada, atribuir o

negar cualidades al objeto evaluado, o simplemente, medir la eficacia de un método o los resultados de una actividad.¹⁹

Recientemente Caraballo R y Rico L en el 2011 plantearon que evaluar es, de manera sucinta, recoger y sistematizar información para tomar decisiones.²⁰

No se puede dejar de considerar la definición recogida en el Reglamento para la Organización del Proceso Docente Educativo en los centros de Educación Médica Superior, Resolución Ministerial No. 15 del 1988. En este se plantea que la evaluación como una parte esencial del proceso de enseñanza – aprendizaje, constituye una vía de retroalimentación para su dirección y para el propio estudiante. Implica el control y la valoración de los conocimientos, habilidades, hábitos, así como de los modos de actuación, que los estudiantes adquieren a través del proceso docente – educativo, al comprobar el grado con que se alcanzan los objetivos propuestos.²¹

La autora de esta investigación asume el concepto emitido en la reciente Resolución Ministerial 210 del 2007 donde se plantea que la evaluación del aprendizaje es un aspecto consustancial al desarrollo del proceso docente educativo. Tiene como propósito comprobar el grado de cumplimiento de los objetivos formulados en los planes y programas de estudio de la educación, mediante la valoración de los conocimientos y habilidades que adquieren y desarrollan los estudiantes; así como, por la conducta que manifiestan en el proceso docente educativo. Constituye, a su vez, una vía para la retroalimentación y la regulación de dicho proceso.²²

A nivel mundial las funciones de la evaluación del aprendizaje ocupan un lugar importante en la literatura pedagógica. Entre estas funciones se describen las siguientes:⁴

Función de retroalimentación: Está directamente vinculada con la dirección del aprendizaje, retroalimenta al profesor de cómo va el proceso docente en correspondencia con los objetivos propuestos a través de diferentes controles y en diferentes momentos. En dependencia de los resultados, el profesor podrá reorientar

su estrategia y acercarse más a los objetivos deseados. Por otra parte le brindará información al estudiante acerca de su progreso en las transformaciones que él debe haber logrado en el sistema de conocimientos y habilidades, y de los valores en sus modos de actuación, le propicia el desarrollo de su autoevaluación y le orienta hacia dónde debe dirigir sus esfuerzos en el trabajo independiente y el autoestudio.

Función instructiva: Contribuye a incrementar, sistematizar, profundizar y generalizar la actividad cognoscitiva, propicia el trabajo independiente y el autoestudio.

Función de comprobación y control: Nos informa sobre el grado en que se cumplen los objetivos de la enseñanza, o sea, nos permite controlar el estado del proceso docente y la efectividad y calidad de la enseñanza y el aprendizaje; de esta manera nos asegura el continuo perfeccionamiento del proceso docente educativo.

Función educativa: Los estudiantes deben hacer en cada control una rendición de cuenta de las responsabilidades que su condición de estudiante les crea ante la sociedad, plantea mayores exigencias y constituye un motor impulsor, consciente y voluntario en el aumento de la calidad del autoestudio y del trabajo independiente, todas sus acciones las convierten de un deber a un placer.

Para que la evaluación cumpla sus funciones, resulta indispensable la participación del estudiante como sujeto activo del aprendizaje, no solo en los momentos aislados, en que se efectúa un control, sino que es necesario concebirla como parte integrante, y a todo lo largo del desarrollo del proceso docente.⁴

La autora comparte el criterio de que estas funciones se relacionan e interactúan entre sí y su dominio y aplicación práctica por parte del profesor posibilita la eliminación de las dificultades y deficiencias que se presentan en la evaluación.

La correcta aplicación de un sistema de evaluación del aprendizaje, que tiene un carácter cualitativo integrador, se basa en un conjunto de condiciones que constituyen las premisas básicas que se deben satisfacer, las cuales se encuentran estrechamente relacionadas entre sí y cuya inobservancia impide que se cumplan efectivamente sus funciones. Estas premisas se estructuran en principios que son:

objetividad, sistematización y continuidad, carácter sistémico, y exigencias como son: la validez y la confiabilidad.⁴

Según el Reglamento para la organización del proceso docente educativo en las universidades médicas en Cuba, la evaluación del aprendizaje puede ser formativa y certificativa.²²

La formativa se desarrolla durante el proceso docente, tiene un carácter diagnóstico y nunca se emplea para emitir un juicio certificativo. En ella se incluyen los controles frecuentes y parciales.

La certificativa o sumativa se desarrolla al concluir una asignatura, ciclo, carrera y especialidad, tiene carácter certificativo e integra la evaluación formativa con el ejercicio evaluativo final que se desarrolla.

Con dicha evaluación se pretende determinar niveles de rendimiento y decidir si se produce éxito o fracaso. Se hace referencia al juicio final global de un proceso que ha terminado y sobre el que se emite una valoración terminal. Su preocupación es poder decir cuánto ha aprendido o progresado el alumno, por eso su finalidad fundamental es la de servir a la selección y jerarquización de alumnos según los resultados alcanzados. Ese carácter terminal suele expresarse en una apreciación cuantitativa del resultado obtenido.²

Para lograr que los controles parciales y finales cumplan con las exigencias de validez y confiabilidad, se requiere una adecuada preparación de los profesores. En este sentido adquiere importancia el trabajo en el departamento docente y en el colectivo de asignatura o disciplina donde la participación del grupo de docentes, preparados para la confección de los instrumentos, asegure la validez y confiabilidad de los mismos.

El estudio y análisis del sistema de evaluación debe constituir una de las líneas principales a desarrollar en el trabajo metodológico que se realiza en los colectivos docentes. En este escenario se precisa el sistema de evaluación a emplear, se determinan las posibles variantes a utilizar y los contenidos objeto de control. Se

define el nivel de profundidad y de asimilación de acuerdo con los objetivos a verificar, se confecciona cada instrumento evaluativo con el empleo de los procedimientos y las técnicas más idóneas para su verificación y se especifican los criterios para su calificación. De esta manera se realiza la planeación de los exámenes.

Planeación del examen.

Con el fin de contribuir a incrementar la efectividad y la eficiencia de los instrumentos evaluativos se proponen en la literatura las etapas siguientes:⁵

1. Definir qué debe ser evaluado.
2. Seleccionar los métodos y procedimientos a emplear.
3. Cómo calificar el examen.
4. Cómo establecer un patrón de aprobado o desaprobado.

Definir qué debe ser evaluado constituye la base de la validez de contenido del instrumento que se confecciona y lo podemos lograr a través de la confección de tablas de contenidos simples y complejos.⁵

En la selección de los métodos y procedimientos a emplear se deben tener en cuenta los objetivos y contenidos a evaluar. Para su elección debemos considerar:⁵

- Emplear con preferencia aquellos que pongan a los educandos en situaciones reales o lo más parecido posible.
- Las tareas clínicas a evaluar prácticamente indican los métodos, procedimientos y técnicas a emplear.
- Valorar las restricciones que pueden condicionar su selección, tales como: claustro docente, tiempo disponible para el examen, cantidad de educandos, recursos para el diseño, entre otros.

En la calificación de los exámenes se utilizan diferentes técnicas que deben estar adecuadamente diseñadas y validadas por un consenso de expertos. Entre las más empleadas encontramos las listas de comprobación y escalas de calificación, claves de calificación de respuestas, observación y grabación de video y las combinadas.⁵

Para la calificación global de un examen se utilizan por lo general tablas de decisiones donde se ponderan las calificaciones de las diferentes preguntas, secciones o problemas ya calificados de forma independiente, ya sea con igual valor o no, con el fin de estandarizar la calificación global del examen.⁵

El establecimiento del patrón de aprobado / desaprobado, es un proceso complejo. Debemos partir del análisis de su contenido en correspondencia con los objetivos educacionales a evaluar. Se han desarrollado diferentes enfoques a partir de la propuesta de Livingston y Ziesky en 1982, citadas por otros autores y que se pueden resumir en dos grandes grupos:⁵

- Enfoque normativo: Establece un patrón relativo que depende de la competencia mostrada por cada examinado. La calificación de aprobado se establece con posterioridad al examen.
- Enfoque criteriado: Parte de un patrón absoluto que se establece con anterioridad a la realización del examen. La selección de la calificación del aprobado es arbitraria e independiente de los resultados que obtengan los estudiantes.

La autora coincide que el método criteriado es el más apropiado para calificar las competencias adquiridas por el estudiante y posibilita una mejor gradación de las calificaciones.

Todos los instrumentos evaluativos deben reunir las características generales siguientes:⁶

- Sus contenidos deben estar directamente relacionados con los objetivos educacionales.
- Ser realista y práctico.
- Tener validez y confiabilidad.
- Ser objetivo y factible de realizar.
- Comprender las cantidades importantes y útiles.
- Ser completo, pero lo más breve posible.
- Ser preciso, pero claro en su redacción.

Los instrumentos evaluativos escritos, según la categoría de las preguntas o ítems pueden dividirse en dos grandes familias: los exámenes de desarrollo, de respuesta libre, tradicionales o de ensayo; los de respuesta estructurada o test objetivos y los mixtos que son productos de la combinación de ambos tipos de preguntas.²³

Por lo antes referido las preguntas o “ítems” pueden ser de respuesta libre o ensayo y de respuesta estructurada o test objetivos.²³

- Respuesta libre o ensayo es aquella en la cual el sujeto aporta su propia respuesta que puede ser una respuesta extensa; corta que no necesita más que una o dos oraciones para responderla; y de completamiento que sólo requiere de una palabra o frase para responderse.
- Respuesta estructurada o test objetivo es aquella en la cual el sujeto elige su respuesta dentro de varias respuestas proporcionadas por el elaborador del examen, tales como la de respuesta alternativa (verdadero o falso); la de selección múltiple y la de igualamiento o pareación, entre otras.

Todos los tipos de ítems, poseen sus ventajas y desventajas; de modo que mientras mejor los conozcamos, aumentan las posibilidades de que podamos utilizarlos mejor, y así elevar sus ventajas y reducir sus desventajas a la mínima expresión.

Las pruebas de tipo ensayo, llamadas también de desarrollo o tradicionales tienen entre sus ventajas:²³

- Miden la capacidad del individuo para relacionar, seleccionar y organizar el material
- Posibilita que el educando exponga sus conocimientos, ordene sus ideas y demuestre su capacidad para expresarse en su idioma.

Entre sus desventajas encontramos:²³

- Limita sensiblemente el número de áreas a explorar
- La objetividad es difícil de conseguir
- Proporciona escasa retroalimentación al educando y al claustro
- Su calificación consume demasiado tiempo, es subjetiva y sus resultados son poco confiables.

Las preguntas de ensayo de respuestas cortas o restringidas, reducen las desventajas de las preguntas de tipo ensayo, se incrementa la muestra de contenido a evaluar y se limitan las preguntas sólo a aquellas que conlleven respuestas libres, pero cortas. Pueden tener forma de pregunta de respuesta breve o de un enunciado con espacios en blanco para llenar o completar. Estas características permiten emplearlas en nuestros exámenes.

Los test objetivos son preferidos por muchos evaluadores por la facilidad, la rapidez y la objetividad de su calificación. La utilización de este tipo de examen escrito tiene las ventajas siguientes:²³

- Asegura objetividad y, en general, propicia el aumento de la confiabilidad y validez del instrumento.
- Incrementa la gama y variedad de los aspectos a evaluar, en poco tiempo y con mínima supervisión.
- Posibilita la facilidad y rapidez de su calificación, lo que permite divulgar con prontitud las calificaciones obtenidas por los educandos.
- Suministra retroalimentación, tanto al profesor como a los educandos, y evidencian los puntos débiles.
- Cuando el número de educandos a evaluar es alto, el costo de su aplicación disminuye.
- El análisis estadístico es relativamente sencillo y suministra información confiable
- Pueden almacenarse bancos de preguntas.

Entre sus desventajas pueden mencionarse:²³

- Resulta difícil al evaluador determinar el grado de profundidad del conocimiento adquirido por el sujeto y en segundo lugar, para cumplir verdaderamente con su cometido requiere de una elaboración muy cuidadosa.
- Su preparación exige mucho tiempo, competencia y práctica con el fin de evitar preguntas ambiguas y arbitrarias. Esta dificultad se incrementa cuando se desea comprobar aspectos complejos de interpretación o solución de problemas, más que los eminentemente factuales.

Calidad de los exámenes certificativos de Pediatría.

- Proporciona a los educandos informaciones e indicaciones que no dispondrían en una situación real.
- Resulta costoso cuando los educandos a examinar son pocos.
- Implica adoptar determinadas condiciones para tener en cuenta las respuestas correctas que pudieran ser adivinadas al azar.
- Exige el entrenamiento previo de los educandos.

Dentro de las preguntas test objetivo más empleadas en los instrumentos escritos en nuestra cátedra encontramos las de selección múltiple tipo complemento simple y agrupado, las de asociación y las de respuesta alternativa.

Las preguntas de selección múltiple son más flexibles y eficaces. Se utilizan para evaluar conocimientos, vocabulario, comprensión, aplicación de principios o habilidad y para interpretar datos. Están compuestas por dos partes fundamentales la guía o núcleo que presenta el problema denominada también base, cuerpo, raíz o tallo y la lista de posibles respuestas o alternativas a la que se le denomina complemento.²³

Entre sus variantes tenemos el tipo complemento simple, donde hay una sola alternativa sencilla para seleccionar dentro de las que se ofrecen, y la de tipo complemento agrupado que ofrece un grupo de aseveraciones y las alternativas que se brindan son combinaciones de estas; son más elaboradas y requieren de procesos mentales más complejos a la hora de seleccionar una alternativa.²³

También dentro de este tipo de pregunta está la de simulación y análisis progresivo que es la más compleja de todas y requiere de un meticuloso esfuerzo en su elaboración. Consiste en la simulación escrita de un caso clínico o una situación de salud a partir de la cual se derivan preguntas de selección múltiple que ganan en complejidad a partir de la situación simulada y que dependen sucesivamente de la o las respuestas elegidas anteriormente.²³

Las preguntas de respuestas alternativas o denominadas de verdadero o falso, constituyen una variante de selección múltiple, con la diferencia de que las alternativas que se emplean son sólo dos. Su utilización debe reducirse a

proposiciones que sean inequívocamente verdaderas o demostrablemente falsas. Se recomiendan cuando no hay suficientes alternativas plausibles para conformar una relación múltiple o que sólo existan dos alternativas posibles en la respuesta.²³

Las de asociación son formas especiales de las preguntas de selección múltiple. Se distinguen de ellas en que en lugar de un solo problema con un grupo de respuestas a escoger, hay varios problemas, cuyas respuestas deben sacarse de una sola lista de alternativas posibles.²³

Juárez Díaz y García Carrasco, en su taller de examen escrito, recogieron la opinión que los instrumentos de opción múltiple son superiores al resto de los exámenes escritos por su capacidad de medir predicciones, discriminación, interpretación, evaluación, extrapolación de los hechos y del conocimiento.²⁴

La autora considera que todos los tipos de incisos o preguntas, mientras que sean bien elaborados y empleados, poseen su utilidad a la hora de confeccionar un instrumento de evaluación. Su estudio constituye interés permanente de los investigadores y de los claustros de profesores en aras de buscar las mejores soluciones para elevar la calidad de los instrumentos que se aplican.

El examen final de la rotación de Pediatría se realiza en el hospital pediátrico de Holguín de forma escrita, su calificación se ejecuta con un enfoque criteriado y se clasifican en aprobados y desaprobados con una escala de 2 a 5, donde resultan desaprobados los que obtienen calificación de 2.

Calidad de los instrumentos escritos.

Tres núcleos conceptuales han sido considerados vitales, por algunos autores, para hablar de calidad de los instrumentos evaluativos: “validez”, “confiabilidad” y “generalizabilidad”.^{4, 25, 26.}

La validez es la correspondencia que existe entre lo que el instrumento evaluativo pretende medir y lo que mide realmente. La definición de validez requiere del criterio de expertos (generalmente externos) y nos capacita para determinar el grado de acierto del diseño del instrumento evaluativo, a partir de los resultados de las

calificaciones de los educandos, sobre la base de una muestra representativa de los mismos.²⁶

Existe confiabilidad cuando se aplica un instrumento evaluativo a un conjunto de personas con la finalidad de medir un determinado conocimiento y se desea tener cierto margen de seguridad con respecto a la posibilidad de obtener resultados similares cuando el instrumento se aplica en varias ocasiones (o bien un instrumento equivalente) a las mismas personas. Es la propiedad de reproducibilidad o estabilidad de las calificaciones que deben cumplir los instrumentos evaluativos.²⁷

La generalizabilidad consiste en estudiar los componentes del error y determinar la contribución de cada uno de ellos en el error total calculado.²⁶

Por lo que se plantea que un instrumento evaluativo tiene calidad en la medida en que es válido y confiable y sus resultados son generalizables.

En la bibliografía revisada aparecen otros conceptos que evalúan la calidad de los instrumentos evaluativos. Entre estos están la objetividad, la pertinencia, el equilibrio, la equidad, la discriminación y la eficacia. Pero son la validez, la confiabilidad y la generalizabilidad las características en la que nos detendremos.²⁶

La Validez usualmente se clasifica en cuatro grupos:²⁶

- a) Validez de contenido.
- b) Validez de criterio.
- c) Validez de construcción.
- d) Validez funcional.

a) Validez de contenido: Está en correspondencia con los conocimientos y las habilidades (tanto cualitativa como cuantitativamente) que evalúan o miden el instrumento de control sean lo suficiente para definir el grado de aprovechamiento del estudiante, sobre la base de los objetivos propuestos.

Esta consiste en el examen sistemático del contenido del instrumento evaluativo para determinar si constituye una muestra representativa del dominio de los

conocimientos y/o habilidades que se pretende evaluar.²⁶

b) Validez de criterio: En este estudio se comparan los resultados de la aplicación del instrumento evaluativo con otros criterios externos que se le llama “variable-criterio”.

Se dice que un instrumento evaluativo tiene validez de criterio cuando, a partir de los resultados obtenidos, se pueden hacer inferencias válidas sobre una variable que no se puede medir directamente por el instrumento evaluativo.

Los estudios de validez de criterio se clasifican en dos tipos:²⁶

- Validez predictiva
- Validez concurrente

La validez predictiva indica la efectividad del instrumento evaluativo en un resultado futuro. Para esto se comparan las calificaciones del instrumento evaluativo con una medida directa de la ejecución posterior de los sujetos llamado “criterio”.

La validez concurrente se refiere a la interrelación del instrumento evaluativo y un criterio de medición efectuado en el momento en que se aplica el instrumento evaluativo.

c) Validez de construcción: Estos estudios se realizan cuando no se dispone de una medida directa del atributo a medir, de tal manera que hay que sustituirlo por un modelo hipotético.

d) Validez funcional: Está en correspondencia con la selección del tipo de actividad (acciones, tareas, problemas) que debemos controlar en el estudiante, para evaluar sus modos de actuación de acuerdo con los objetivos propuestos. Ambos tipos de validez se interrelacionan y complementan.²⁶

Es una variante de la validez de contenido, pero que adquiere personalidad propia en la educación médica. Está en correspondencia con el tipo de actividad con que debemos verificar el nivel alcanzado por el educando en un grupo de habilidades que conforman su modo de actuación.²⁶

En la medida en que las actividades de la evaluación y el diseño del instrumento se aproximan a la realidad, o sea, estén dirigidas a valorar o medir la aplicación social e independiente del sistema de conocimientos, hábitos, habilidades y modos de actuación por el estudiante, mayor será la validez de dicho control. Mientras más teórico, reproductivo y alejado de la práctica esté un control mucho menor será su validez.

La validez y la confiabilidad están estrechamente relacionadas. Un control que cumpla con las exigencias de la validez, tiene un alto grado de probabilidad de ser confiable; sin embargo, no necesariamente ocurre así a la inversa. Por ejemplo, un control puede ser confiable, pues al ser calificado por diferentes profesores se obtienen los mismos resultados y sin embargo, no tener validez ya que el instrumento no mide los objetivos establecidos. Es por ello que en primer orden hay que garantizar la validez de todo instrumento evaluativo y complementar en segundo orden su confiabilidad.

El profesor Díaz Rojas²⁸ propone una metodología para medir la calidad de un instrumento escrito, en la que sugiere al índice de dificultad, al índice de discriminación y al coeficiente de discriminación del punto biserial como indicadores para el análisis de la validez de los exámenes y al coeficiente alfa de Cronbach para medir su confiabilidad. Indicadores que fueron utilizados por la autora en la investigación.

Wood citado por Backhoff Escudero²⁹ define como nivel de dificultad de un ítem a la proporción de estudiantes que responden correctamente un inciso de un examen. Cuanto mayor sea esta proporción, menor será su dificultad, por lo que se trata de una relación inversa: a mayor dificultad del ítem, menor será su índice.

El índice de dificultad del ítems queda determinado por el número de examinados en el grupo que lo resuelven correctamente. El índice de dificultad de los ítems oscila entre cero y uno. Si el índice se ubica cercano a cero, refleja un alto grado de dificultad, pero si por el contrario se ubica cercano a uno, entonces su grado de dificultad es bajo (ítem fácil), es decir su resolución no produce dificultad al

estudiante para su contestación. Lo más apropiado es que el índice de dificultad esté próximo a 0.50.³⁰

El manual del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) de la universidad de Baja California en México considera el nivel medio de dificultad del examen entre 0.5 y 0.6.²⁹

El índice de dificultad permite determinar en qué medida una pregunta es fácil o difícil. En realidad es un índice de facilidad, puesto que cuanto más alto es, más fácil es la pregunta. Este índice va de 0 muy fácil a 100 muy difícil. El índice de dificultad recomendado se sitúa entre 50 y 60.³⁰

La autora Estela Diamant del Ministerio de Cultura y Educación de la República Argentina define a la Dificultad como la característica del comportamiento de un ítem que varía en forma inversa con el porcentaje de casos de una muestra que respondió correctamente. Esto quiere decir a menor porcentaje de aciertos se considera mayor dificultad del ítem. Se denomina índice de dificultad a la proporción de individuos de un grupo específico que respondieron incorrectamente al ítem.³¹

Un inciso puede ser muy fácil porque las opciones con función distractora son inverosímiles o el contenido explorado puede estar en un nivel reproductivo. Puede ser muy difícil debido a que la opción correcta es errónea, existen dos o más opciones correctas, el contenido explorado es raro, rebuscado o minucioso o el enunciado no está claramente definido.²⁶

El Índice de dificultad en sí mismo es de utilidad en el análisis de los resultados de un instrumento evaluativo, pero una mayor utilidad reporta el análisis de la correspondencia entre el Índice de dificultad esperado o pronóstico y el obtenido o real del examen. El Índice de dificultad pronóstico se determina a partir del estándar propuesto por Backhoff²⁹ para el análisis de dificultad en instrumentos de evaluación, el que establece que dichos instrumentos deben tener un:

- 5% de incisos fáciles.
- 20% de incisos medianamente fáciles.

Calidad de los exámenes certificativos de Pediatría.

- 50% de incisos de dificultad media.
- 20% de incisos medianamente difíciles.
- 5% de incisos difíciles.

Según el criterio de otros autores una prueba que posea dificultad media es superior en confiabilidad de resultados tanto a las que tienen mucha dificultad como a las que resultan muy fáciles. Un buen examen será aquel que sea superado por algo más de la mitad de los educandos.³²

El poder de discriminación del examen no es más que la cualidad de cada pregunta del instrumento que permite distinguir los estudiantes con altos y bajos rendimientos y aporta criterio de predicción sobre sus resultados pues si el examen y una pregunta miden la misma habilidad o competencia, se puede esperar que quien tuviera una puntuación alta en todo el examen deberá tener altas probabilidades de contestar correctamente los temas evaluados en sus preguntas.^{9, 33.}

Usualmente se utilizan dos formas para medir la capacidad discriminativa del ítem: el índice de discriminación y el coeficiente de discriminación.

El índice de discriminación expresa la medida en que el ítem o pregunta discrimina entre los buenos y peores estudiantes. Se entiende que los buenos estudiantes deben responder en su mayoría correctamente el ítem y los malos incorrectamente. Nos permite predecir el comportamiento del ítem o pregunta dentro del examen, para la elaboración de nuevos instrumentos con estructura bastante similar. Entre más alto es el índice de discriminación, el ítem diferenciará mejor a los estudiantes con altas y bajas calificaciones.³²

El máximo valor del índice de discriminación de un ítem tiene estrecha relación con su índice de dificultad, ya que un reactivo con valor muy alto o muy bajo de índice de dificultad en realidad no discrimina, pues es solucionado por la mayoría de los estudiantes o por la minoría de ellos respectivamente.²⁸

Se propone en la metodología revisada utilizar la norma discriminativa, que permite determinar el valor óptimo que debe presentar un ítem en su índice de

discriminación, de acuerdo al índice de dificultad del mismo.²⁸

Con la norma discriminativa se puede elaborar otro indicador que se conoce como relación discriminativa y está dado por el cociente entre el índice de discriminación del ítem y su norma discriminativa, esta nos permita saber si un reactivo es aceptable o no según su valor de índice de discriminación.²⁸

En la Facultad de Medicina de la Universidad autónoma de México, consideran un índice de discriminación máximo cuando el valor que se obtiene es de uno, alto cuando el índice obtenido oscila entre 0.41 y más, moderado cuando se encuentran entre 0.21 y 0.40, bajo de 0.01 a 0.20 y negativo cuando el índice es igual a cero.²⁴

En esta institución sostienen que la distribución ideal de los diversos grados de discriminación son: alto más de un 25 % de reactivos del examen, moderado entre 24 y 16 por ciento, bajo menor del 15 % y negativo menos del 5 % del examen.²⁴

El coeficiente de discriminación se mide a través de dos formas: el punto de correlación biserial y el coeficiente de correlación biserial. La ventaja de utilizar el coeficiente de discriminación en lugar del índice de discriminación es que con el primer método se toman en cuenta todas y cada una de las personas evaluadas, mientras que con el segundo, sólo se toma 54 % de ellas (27 % más alto y 27 % más bajo).²⁸

El punto de correlación biserial se mide para saber si los estudiantes de alto rendimiento son los que obtienen las respuestas adecuadas mientras que el coeficiente de correlación biserial (rbis) es para determinar el grado en que las competencias que mide el test también las mide el reactivo.²⁸

En la literatura especializada se reportan varios procedimientos para cuantificar la confiabilidad de los instrumentos evaluativos. Entre ellos se citan:³⁴

- El coeficiente θ .
- El coeficiente ω (El coeficiente ∞)
- El coeficiente de Kuder-Richarson.
- El coeficiente α de Cronbach.

Según criterios de varios autores y compartidos por la autora, el coeficiente α de Cronbach.^{30, 33, 34.}

- Es el más utilizado en los estudios contemporáneos de calidad de los instrumentos evaluativos, lo que nos permite comparar nuestros resultados con los que se reportan en la literatura mundial.
- Nos proporciona una medida de la consistencia interna de los instrumentos evaluativos.
- Puede calcularse aún en el caso de que empleemos preguntas de diferentes formatos en el instrumento evaluativo.

Múltiples han sido los reportes con respecto al valor mínimo que debe alcanzar el coeficiente α de Cronbach para declarar como confiable a un instrumento evaluativo. Pero la autora considera que se debe actuar con cautela a la hora de la decisión de este punto de cohorte.

Algunos ejemplos de estos reportes son: en los exámenes de Simulación Oral en el Consultorio Médico (SOOs) donde se han obtenido coeficientes entre 0,54 y 0,63³⁵ y consideran como aceptables los superiores a 0,60; Anastasi³⁶ y Weiner³⁷ aceptan como adecuados los coeficientes que oscilan entre 0,80 y 0,85; el Colegio de Médicos de Familia de Canadá en sus exámenes de Certificación de especialistas reporta para los exámenes escritos basados en respuestas cortas como aceptables, coeficientes entre 0,69 y 0,71³⁶; en los tipo Examen Clínico Objetivo Estructurado (ECO) con pacientes estandarizados y preguntas de ensayo de respuestas cortas han considerado como aceptables coeficientes entre 0,71 y 0,77.^{38, 39.} En este tipo de examen otros autores han reportado coeficientes entre 0,40 y 0,88.⁴⁰

En Cuba a partir de los exámenes que se diseñan y aplican desde el año 1996, en el marco de las inspecciones territoriales del Ministerio, se han realizado varios estudios y la Comisión Nacional de Evaluación de la Competencia y el Desempeño Profesional decidió establecer como aceptable, un coeficiente alfa de Cronbach de 0,6.³⁴ La autora tomó este valor como referencia para evaluar dicho indicador en su investigación.

Calidad de los exámenes certificativos de Pediatría.

A partir de todo lo expuesto la autora asume en este estudio el uso de los indicadores de calidad: índice de dificultad, índice de discriminación, coeficiente de correlación punto biserial y coeficiente alfa de Cronbach para evaluar la calidad de los exámenes certificativos y continuar el perfeccionamiento de la planeación de los instrumentos evaluativos en la cátedra de Pediatría.

OBJETIVOS

General.

Caracterizar la calidad de los exámenes certificativos de Pediatría aplicados a estudiantes del cuarto año de la carrera de medicina del municipio Holguín, en el curso 2012 – 2013.

Específicos.

1. Describir las características de los exámenes según estructura, temas y áreas de aprendizaje exploradas y los tipos de preguntas e incisos utilizados.
2. Identificar la correspondencia entre los contenidos de la unidad curricular y su exploración en los exámenes.
3. Valorar el comportamiento del índice de dificultad, índice y coeficiente de discriminación del punto de correlación biserial y el coeficiente alfa de Cronbach.

DISEÑO METODOLÓGICO

Se realizó un estudio de evaluación, en el campo del proceso enseñanza-aprendizaje en el área de la evaluación, cuyo objeto de análisis fue la calidad de los exámenes certificativos aplicados en el cuarto año de la carrera de medicina del municipio Holguín en el curso 2012-2013.

El **universo** y la **muestra** estuvieron constituidos por la totalidad de los exámenes ordinarios de Pediatría del primer semestre (Examen I) y segundo semestre (Examen II) de los estudiantes que cursaron el cuarto año de la carrera de Medicina del municipio Holguín durante el curso 2012 -2013.

Para el desarrollo de la investigación se utilizaron métodos teóricos, empíricos y procedimientos estadísticos.

Métodos Teóricos.

El análisis documental, para la revisión bibliográfica de temas relacionados con el proceso enseñanza aprendizaje, evaluación, planeación de los exámenes, calidad de los instrumentos de evaluación, documentos como el Programa Nacional de Pediatría, Programa de la Carrera de Medicina, exámenes de la disciplina y otros documentos normativos y reguladores del proceso docente.

El análisis síntesis en el procesamiento de la información teórica y empírica, al realizar el análisis de los exámenes, la elaboración de los resultados y conclusiones sobre la calidad de los exámenes aplicados

El de inducción deducción para arribar a generalizaciones acerca de las potencialidades y limitaciones que presentan los instrumentos de evaluación aplicados.

Métodos Empíricos.

Utilizamos la consulta a especialistas y grupos focales coordinado por la autora. Se realizó un grupo focal con profesores y otro con estudiantes al concluir la evaluación de la asignatura en el segundo semestre del curso.

Calidad de los exámenes certificativos de Pediatría.

La consulta a especialistas se hizo con pediatras de más de 20 años de experiencia docente y se utilizó para clasificar los exámenes según tipo de preguntas e incisos que conforman los instrumentos, permitió determinar los criterios para definir los intervalos de clasificación del índice de dificultad y clasificar los incisos e ítems según los contenidos de la unidad curricular y área de aprendizaje que se exploraron. (Anexo 1)

El grupo focal de profesores se realizó con profesores especialistas en Pediatría, con una experiencia docente de más de 15 años. Se persiguió como objetivo obtener información en relación a la calidad de los exámenes ordinarios de Pediatría del curso 2012 – 2013. Se efectuó en dos sesiones de trabajo una con los de la atención primaria y otra con la atención secundaria. (Anexo 2)

Para el grupo focal con estudiantes se seleccionó al azar el 20 por ciento de cada uno de los 5 grupos existentes. Persiguió igual objetivo que el grupo focal con profesores. El objetivo también fue obtener información en relación con la calidad de los exámenes ordinarios de Pediatría del curso 2012 – 2013. Se llevó a cabo en tres sesiones de trabajo, dos con los que rotan por el hospital y una con los que rotan por los policlínicos. (Anexo 3)

El análisis de la información obtenida en dichas entrevistas fue cualitativo, se resaltaron los elementos en los que hubo consenso y en los que hubo discrepancia.

Obtención de los datos.

Se utilizaron como datos los obtenidos de la revisión directa de los incisos y preguntas de los dos exámenes ordinarios de Pediatría del curso 2012-2013, el del primer semestre (Examen I) y del segundo semestre (Examen II).

Se utilizó la hoja de cálculo de Microsoft Excel donde se tabularon los resultados de los exámenes. Se incluyeron los resultados de cada inciso categorizados en correcto ® e incorrecto (I), de cada pregunta y la nota final según la calificación obtenida de ambos instrumentos.

Procesamiento de los datos.

I. Descripción de las características de los exámenes.

Se caracterizaron los exámenes y se tuvieron en cuenta:

- ✓ Temas de la asignatura, se obtuvieron del programa y del plan calendario docente.
- ✓ Tiempo en que se impartió el tema de la asignatura, se obtuvo del Plan calendario.
- ✓ Estructura de los exámenes según cantidad de preguntas e incisos, se obtuvieron de los exámenes.
- ✓ Clasificación de las preguntas e incisos de los exámenes según forma en que fueron elaboradas.
- ✓ Clasificación de los incisos de acuerdo al contenido temático que exploraron.

II. Correspondencia entre los contenidos de la unidad curricular y su exploración en los exámenes.

Se calculó el número de horas que se le dedica a cada contenido de la unidad curricular según el plan calendario de la asignatura y se determinó el por ciento que representa del total de horas dedicadas a esta estancia.

Se clasificaron los incisos de los exámenes según el contenido temático que exploraron, se determinó el por ciento de incisos que exploraron cada temática en base al total de incisos o ítems de cada examen.

Se utilizó el coeficiente de correlación de Spearman para identificar si existe correspondencia entre el tiempo dedicado en el plan calendario a los contenidos de la unidad curricular y su representación en los incisos de los exámenes. Los valores del coeficiente de correlación (r) próximos a 1 significan que la distribución de las preguntas en los exámenes se realizó en correspondencia con el tiempo que se le dedica a cada temática en el plan calendario de la asignatura.

III. Indicadores de Calidad.

3.1. Determinación del índice de dificultad para los incisos, temáticas, preguntas y exámenes.

Calidad de los exámenes certificativos de Pediatría.

A partir de los datos primarios que obtuvimos de la revisión de cada examen ordinario se procedió al cálculo del índice de dificultad para cada inciso, pregunta, temática y examen.

Para identificar el índice de dificultad de un inciso, se dividió el número de estudiantes que contestaron correctamente el ítem entre el número total que contestaron el ítem. Esta proporción se le denota como D_i y se utilizó la siguiente fórmula:²⁸

$$D_i = \frac{A}{N}$$

Donde:

D_i = Índice de dificultad del ítem.

A = Número de respuestas correctas del ítem.

N = Número total de estudiantes que contestaron el ítem.

Cuanto mayor sea esta proporción, menor será su dificultad. Lo que quiere decir que se trata de una relación inversa: a mayor dificultad del ítem, menor será su índice.

Para identificar el índice de dificultad de una pregunta, se dividió el número de estudiantes que contestaron correctamente la pregunta entre el número total de estudiantes que se presentaron al examen, esta relación se denomina con la letra D_p .

$$D_p = \frac{\sum (A_{p1} + A_{p2} + \dots + A_{pp})}{\frac{n_{ip}}{Nt}}$$

Donde:

D_p = Índice de dificultad de la pregunta.

A_p = Número de respuestas correctas de los incisos de la pregunta.

N_{tp} = Número total de incisos de la pregunta.

N_t = Número total de presentados en el examen.

Para calcular el índice de dificultad de cada temática explorada en el examen, se

sumó el total de estudiantes que respondieron correctamente cada uno de los incisos que corresponden a la temática explorada y se dividió entre el número de incisos que exploró esa temática en el examen, obteniéndose la media de respuestas correctas para la temática; luego se calculó el índice de dificultad el cual se corresponde con el promedio de respuestas correctas entre el total de examinados, como se representa en la siguiente fórmula:²⁸

$$Dt = \frac{\sum (A_{t1} + A_{t2} + \dots + A_{tn})}{\frac{n_{tt}}{Nt}}$$

Donde:

Dt = Índice de dificultad de la temática.

At = Número de respuestas correctas de los incisos de la temática.

Ntt = Número total de incisos que exploran la temática.

Nt = Número total de presentados en el examen.

➤ **Determinación de los intervalos para la clasificación de los incisos según su Índice de dificultad.**

Se realizó sobre la base del consenso del criterio de especialistas, y se definieron los intervalos según la proporción de estudiantes aprobados en ellos, se estableció que los incisos que tuvieron un Índice de dificultad.²⁸

- Menor de 0.40, se consideran difíciles.
- Entre 0.41 y 0.50, se consideran medianamente difíciles.
- Entre 0.51 y 0.80, se consideran de dificultad media.
- Entre 0.81 y 0.90, se consideran medianamente fáciles.
- Entre 0.91 y 1, se consideran fáciles.

➤ **Distribución de los incisos según el índice de dificultad pronóstico y el obtenido en cada examen.**

Se distribuyeron los incisos de los exámenes según su índice de dificultad obtenido en cada uno de los intervalos de clasificación definidos previamente.

El Índice de dificultad pronóstico se determinó a partir del estándar propuesto por Backhoff²⁹ para el análisis de dificultad en instrumentos de evaluación, el que establece que dichos instrumentos deben tener un:

- 5% de incisos fáciles.
- 20% de incisos medianamente fáciles.
- 50% de incisos de dificultad media.
- 20% de incisos medianamente difíciles.
- 5% de incisos difíciles.

Se elaboró una tabla de doble entrada en cuyas columnas se plasmaron las frecuencias absolutas y relativas de los índices de dificultad esperados y reales de cada examen, se utilizó la prueba estadística de correlación de rangos de Spearman para determinar la correspondencia entre lo esperado y lo real.

3.2. Determinación del poder de discriminación a través del índice y el coeficiente de discriminación para cada inciso y pregunta.

Para determinar el índice de discriminación se utilizó la siguiente fórmula:²⁸

$$D_i = \frac{GA_{aciertos} - GB_{aciertos}}{N_{grupo\ mayor}}$$

Donde:

D_i = Índice de discriminación del reactivo i

$GA_{aciertos}$ = Número de aciertos en el reactivo i del 27% de personas con las puntuaciones más altas en el test.

$GB_{aciertos}$ = Número de aciertos en el reactivo i del 27% de personas con las puntuaciones más bajas en el test.

$N_{grupo\ mayor}$ = Número de personas en el grupo más numeroso (GA o GB).

Mientras más alto es el índice de discriminación, el reactivo diferenciará mejor a las personas con altas y bajas calificaciones. Si todas las personas del GA contestan correctamente un reactivo y todas las personas del GB contestan incorrectamente, entonces $D= 1$ (valor máximo de este indicador); si sucede lo contrario, $D= -1$ (valor

Calidad de los exámenes certificativos de Pediatría.

máximo negativo); si ambos grupos contestan por igual, $D= 0$ (valor mínimo de discriminación).²⁸

Algunos autores definieron la siguiente regla para determinar la calidad de los reactivos, en términos del índice de discriminación.

D	Calidad	Recomendación
> 0.39	Excelente	Conservar
0.30 – 0.39	Buena	Posibilidades de mejorar
0.20 – 0.29	Regular	Necesidad de revisar
0.00 – 0.20	Pobre	Descartar o revisar a profundidad
< -0.01	Pésima	Descartar definitivamente

Para determinar el valor óptimo que debe presentar un ítem en su índice de discriminación de acuerdo al índice de dificultad se utilizó la norma discriminativa cuyo valor lo obtuvimos con la siguiente fórmula

$$ND = \begin{cases} 0,3 * ID & \text{si } 27 \% \leq ID \leq 73 \% \\ 100 - ID & \text{si } 73 \% \leq ID \leq 100 \% \end{cases}$$

A partir de la norma discriminativa se determinó la relación discriminativa que indica si un reactivo es aceptable o no según su valor de índice de discriminación, y está dado por el cociente entre el índice de discriminación (D) del ítem y su norma discriminativa (ND):²⁸

$$RD_i = \frac{D_i}{ND_i}$$

El coeficiente de discriminación punto de correlación biserial se utiliza para saber si los estudiantes de más alto rendimiento obtienen las respuestas correctas, es decir que tanto valor predictivo tiene la pregunta e inciso y como se puede contribuir a las predicciones. Se calculó con la siguiente fórmula:²⁸

$$r_{pbis} = \frac{\overline{X_1} - \overline{X_0}}{S_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

Donde:

X₁= Media de las puntuaciones totales de aquellos que respondieron correctamente el ítem.

X₀= Media de las puntuaciones totales de aquellos que respondieron incorrectamente el ítem.

S_x = Desviación estándar de las puntuaciones totales.

N₁= Número de casos que respondieron correctamente el ítem.

N₀= Número de casos que respondieron incorrectamente el ítem.

N = n₁+ n₀

Según estándares internacionales la distribución de los resultados del coeficiente punto de correlación biserial es:

r pbis < 0 preguntas que discriminan negativamente.

R pbis = 0 – 0.14 preguntas que discriminan pobremente.

R pbis = 0.15 – 0.25 preguntas que discriminan regular.

R pbis = 0.26 – 0.35 preguntas con buen poder discriminativo.

R pbis > 0.35 preguntas con excelente poder de discriminación.

3.3. Determinación del coeficiente alfa de Cronbach.

El coeficiente alfa de Cronbach permite determinar la consistencia interna de un instrumento y brinda información relacionada con el aporte que hace cada pregunta a la consistencia general del examen. Se calculó con la siguiente fórmula:²⁸

$$\alpha = \frac{nr}{1 + r(n-1)}$$

Donde n es el número de elementos del examen y r su correlación media.

Para su interpretación se toma en cuenta cómo aumenta o disminuye el alfa sin la pregunta, si el valor es igual o inferior al alfa del examen la pregunta aporta

consistencia interna al examen, pues al eliminarla el coeficiente desciende; en caso contrario, si el valor del alfa del examen se eleva al eliminar la pregunta, es un indicador de que la pregunta afecta la consistencia interna y disocia los resultados del examen.

Procedimientos estadísticos

Se utilizaron, el análisis de frecuencias absolutas y relativas, el coeficiente de correlación de Spearman para validar la correspondencia entre el fondo de tiempo dedicado a los diferentes contenidos en el plan calendario de las asignaturas y la frecuencia en que se exploraron en los exámenes y para validar la correspondencia entre los resultados de los valores del índice de dificultad esperados y los observados, el coeficiente alfa de Cronbach para conocer la consistencia interna de los exámenes.

Finalmente los resultados se triangularon a los efectos de dar salida a los objetivos de la investigación, a través de los métodos empíricos, teóricos y procesamiento estadístico de los resultados.

En cuanto a los aspectos éticos de la investigación, se consideró la voluntariedad de participación de los implicados a los cuales se les comunicó que las informaciones obtenidas solo serían utilizadas con fines científicos de igual forma se garantizaron las condiciones para desarrollar las sesiones de trabajo.

En la elaboración de este trabajo se utilizó una computadora Dual Core con ambiente de Microsoft Windows 7. Para el almacenamiento y procesamiento de los datos primarios, cálculo de porcentaje e indicadores de calidad se utilizó la aplicación Microsoft Excel; el Microsoft Word para la redacción del informe final y el Microsoft Power Point para la exposición. Para los resultados estadísticos fue utilizado el paquete estadístico profesional MyStat v. 12.0 para Windows. Los resultados se mostraron en cuadros y gráficos.

ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

A. Resultados de la aplicación de los instrumentos y técnicas colectivas.

El grupo focal con profesores se realizó en dos secciones de trabajo, una con ocho profesores de atención secundaria y otra con ocho de atención primaria. Se realizó tres secciones con estudiantes, dos de 15 estudiantes cada una que rotaron por el hospital y otra de 16 que rotaron por los policlínicos.

Todos estos encuentros coordinados por la autora para la discusión de la guía elaborada al efecto. En la discusión hubo consenso en la totalidad de los participantes tanto de profesores como estudiantes. Las principales conclusiones a las que arribaron fueron:

Profesores.

- Existe correspondencia entre los problemas de salud que se abordan en los exámenes con los que se enfrenta el médico general.
- Se exploró un número elevado de problemas de salud en correspondencia con el número de temas del plan calendario de la asignatura como consecuencia de la planeación de los instrumentos.
- Se sugiere explorar más la promoción y prevención de las enfermedades por ser los niveles de atención en que más deben trabajar los médicos generales.
- En ambos exámenes existen preguntas en las que se exploran gran número de temáticas y contenidos sin delimitar su área lo que puede traer confusión al estudiante.
- El examen I tiene un carácter reproductivo que permite clasificar a la mayoría de las preguntas como medianamente fáciles.
- El examen I carece de preguntas de ensayo en la que se pueda evaluar la instrucción 1/2009 referente a la ortografía.
- En la pregunta 7 del examen I clasificada como de ensayo de respuestas corta tipo completamiento detectaron deficiencias en su confección por encontrar artículos delante del espacio a completar que puede sugerir la respuesta; sin embargo esta deficiencia metodológica de la pregunta no interfirió en los resultados y se clasificó como de dificultad media.

Calidad de los exámenes certificativos de Pediatría.

- El ítem 7.2 del examen I admite varias respuestas que no son contempladas en la clave de calificación o en su redacción debe especificar la característica del contenido de los vómitos que no debe faltar, esta deficiencia apoya el resultado del procesamiento estadístico de este ítem y de la pregunta en general.
- Los resultados de los ítems 7.4 y 7.6 del examen I que responden al tema de enfermedades cardiovasculares no fueron alentadores y evidencian falta de preparación de los estudiantes.
- El examen II cuenta con mayor número de preguntas con nivel de asimilación productivo que permite clasificar al examen de dificultad media.
- La pregunta 3 del examen II, es una pregunta de ensayo que se orientó en forma de cuadro y donde se especificó las características de los síntomas principales a diferenciar; evidenció falta de interpretación y conocimiento de los estudiantes y resultó la temática más difícil del examen.
- En la pregunta 6 el ítem 6.8 y 6.9 trata del contenido de la sepsis, muy presente en la morbilidad y mortalidad actual en el niño, y por los resultados estadísticos vemos que los estudiantes no lograron vencer el objetivo del diagnóstico y la conducta.

Estudiantes.

- Los problemas de salud que se abordaron en los exámenes fueron impartidos en las diferentes rotaciones.
- Los que rotan en la atención primaria, los seminarios que realizan no se corresponden con el contenido que han recibido en ese momento, eso sucede cuando están en las rotaciones de una semana de miscelánea I y II que son especialidades. El profesor de la atención primaria espera a que todos los estudiantes del grupo hayan rotado por todas las especialidades de esa rotación y luego hacen el seminario y la clase taller todo el grupo junto.
- Durante la rotación por Gastro recibieron todos los contenidos y las evaluaciones frecuentes los prepararon para el examen final; al culminar la rotación aplican un examen que los prepara para enfrentarse a este tipo de preguntas al finalizar la estancia por Pediatría.

Calidad de los exámenes certificativos de Pediatría.

- Los exámenes intrasemestrales los familiarizan con los tipos de preguntas para lograr enfrentar el examen final.
- Las evaluaciones frecuentes en los diferentes servicios y en las áreas de salud se realizan en su mayoría con situaciones problemáticas.
- Los problemas de salud explorados en los exámenes son entidades frecuentes a las que se enfrentan una vez graduados y les permiten su adecuado desempeño como médicos generales.
- Existencia de preguntas en los exámenes que exploran varias temáticas sin delimitarlas, lo cual trae dificultades para la comprensión y elección de la respuesta correcta.
- La pregunta 3 del examen II, relacionada con las afecciones quirúrgicas, la encontraron compleja, pero refirieron que les faltó estudio pues es un contenido muy bien impartido por los profesores de esa rotación.
- La pregunta 4 del examen II, refirieron estuvo compleja pues estaba formada por ítems de diferente formato y algunas respuestas dependieron de la anterior pero coinciden que los conocimientos recibidos en la rotación por Gastro les permitió responderla.
- El contenido del tema sepsis e inmunidad opinan se imparte en corto tiempo solo una semana y son temas complejos para su comprensión.
- En la pregunta 6 el ítem 6.8 y 6.9 que exploró el diagnóstico de la sepsis y su conducta reconocieron no saber diferenciar bien entre sepsis, bacteriemia y síndrome de respuesta inflamatoria sistémica, unos por falta de estudio y otros refieren no haberlo comprendido en la práctica.

En ambos grupos focales los criterios de coincidencia son:

- ❖ Correspondencia entre los problemas de salud que se abordan en los exámenes y los problemas de salud a los que se enfrenta el médico general.
- ❖ Presencia de preguntas en los exámenes que exploran gran número de temáticas y contenidos sin delimitar su área lo que trae confusión al estudiante.
- ❖ La pregunta 3 del examen II, fue compleja para los estudiantes y evidenció falta de estudio.

- ❖ La mayoría de los estudiantes no lograron vencer el objetivo del diagnóstico y la conducta de la sepsis que se debió a la falta de estudio y pobre comprensión del tema.

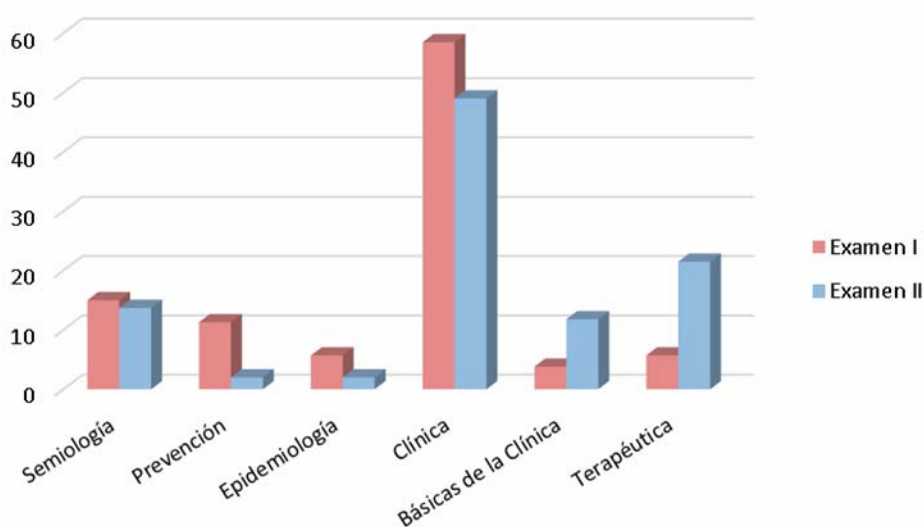
B. Resultados del análisis de calidad de los instrumentos.

I. Características de los exámenes.

Ambos exámenes están compuestos por 7 preguntas, el examen I con 53 incisos y el II con un total de 51. En ambos instrumentos se exploraron la totalidad de las temáticas como son temas generales, nutrición, inmunidad y sepsis, recién nacido, enfermedades digestivas/desequilibrio hidromineral y ácido básico, enfermedades respiratorias, del sistema endocrino, del sistema hemolinfopoyético y cáncer, del aparato cardiovascular, renales, del sistema nervioso y afecciones quirúrgicas más frecuentes.

Son exámenes mixtos con preguntas de desarrollo y de test objetivo, el examen I aplicado a 192 estudiantes y el II a un total de 202. Las áreas de aprendizaje que estuvieron representadas en ambos exámenes fueron la semiología, prevención, epidemiología, clínica, básicas de la clínica y terapéutica. La frecuencia de incisos con que se exploró cada área en el examen aparece representada en el gráfico 1.

Gráfico 1. Por ciento de incisos por área de aprendizaje que explora cada examen.



Calidad de los exámenes certificativos de Pediatría.

El área clínica fue la más explorada en los dos exámenes con 58.5 y 49 porcientos en correspondencia con los objetivos del programa de la asignatura.

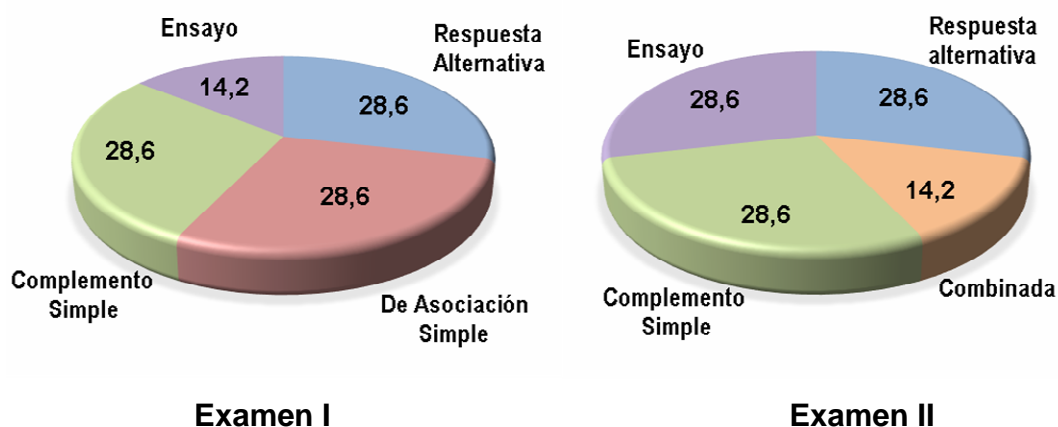
Los especialistas entrevistados son de la opinión que es en el ciclo clínico de la carrera donde el estudiante debe vencer el objetivo del diagnóstico de las enfermedades más frecuentes con el empleo del método clínico como forma de investigación lo que justifica el incremento en los exámenes de esta área de aprendizaje.

En el ciclo básico de la carrera de medicina los estudiantes reciben contenidos relacionados con la promoción y prevención de salud de los individuos, que incluye a los niños, así como para las familias y comunidad. En los exámenes analizados se retomaron algunos de estos contenidos y se pone de manifiesto la interdisciplinaridad con la asignatura rectora.

En el examen II, es de señalar que esta área de aprendizaje es poco explorada al igual que la epidemiología, ambas de vital importancia para que el egresado trabaje en su área de salud con enfoque de riesgo. También en este mismo examen el área terapéutica está muy indagada con más elementos farmacológicos que preventivos.

Dentro de los objetivos instructivos del programa está la realización de acciones de promoción, prevención, curación y rehabilitación encaminadas a lograr el incremento del estado de salud de la población pediátrica, la familia y la comunidad. El grupo de profesores coinciden con la autora que debemos incrementar las preguntas relacionadas con las acciones de promoción y prevención por ser las más relacionadas con los modos de actuación del futuro profesional.

Gráfico 2. Por ciento del tipo de pregunta para cada examen de Pediatría.



En el gráfico 2 se visualiza el porcentaje del tipo de preguntas para cada examen. En ambos predominaron las de test objetivo y dentro de ellas las de selección múltiple tipo complemento simple con 28.6 %, de respuesta alternativa con igual por ciento y en el examen I las de asociación simple también con 28.6 %. Estas preguntas son muy ventajosas cuando se trata de un número elevado de estudiantes a evaluar pues sus calificaciones son más fáciles y permiten medir varios contenidos esenciales o básicos.

En comparación con el examen I en el II se planificó más preguntas de ensayo con 28.6 %, una de ellas de respuesta más extensa. Este tipo de preguntas permite que el estudiante exponga sus conocimientos, ordene sus ideas y demuestre su capacidad de expresión; así como una mayor evaluación de la resolución ministerial 1/2009 del Ministerio de Educación Superior relacionada con los errores ortográficos, que hoy constituyen una dificultad real dentro del grupo estudiantil.

En este segundo examen encontramos una pregunta combinada, la número cuatro, que exploró la temática de enfermedades digestivas. En ella se utilizaron ítems de test objetivo tipo simulación y análisis progresivo y otros de ensayo a punto de partida de una situación problémica. Esta pregunta, a criterio de los estudiantes, estuvo compleja pero muy bien elaborada lo cual coincide con lo planteado por los profesores en el grupo focal y con la autora.

López Espinosa y colaboradores en un estudio realizado en la Universidad de Ciencias Médicas de Villa Clara, revisaron un grupo de exámenes de las asignaturas que conforman la disciplina rectora correspondiente a las carreras de Medicina, Estomatología y Enfermería. En dichos instrumentos se constató la pobre utilización de preguntas de simulación y análisis progresivo por su complejidad y meticuloso esfuerzo en su elaboración.⁴¹

Examen I.

La pregunta uno estuvo conformada por 9 incisos de respuesta alternativa y exploró contenidos de tres temas con tres incisos cada uno. Dentro de los Temas Generales se trató el desarrollo psicomotor, la evaluación del estado nutricional y las mensuraciones. En la Nutrición se indagó sobre lactancia materna, ablactación y desnutrición proteico energética por defecto. En el tema de Sepsis e inmunidad se trataron las vacunas, el desarrollo del sistema inmune y la profilaxis de la sepsis. Las áreas de aprendizaje exploradas fueron semiología, clínica y prevención.

La pregunta dos tuvo 8 incisos de asociación simple donde el estudiante debió enlazar según correspondía los síntomas y signos con diferentes enfermedades del sistema respiratorio. La clínica esta vez fue el área examinada.

La pregunta tres fue de selección múltiple tipo complemento simple, la cual exploró contenidos de las enfermedades diarreicas agudas, persistente, desequilibrio hidroelectrolítico y ácido básico y parasitismo intestinal. En ella se indagaron las áreas de la clínica, las básicas de la clínica, prevención y terapéutica.

La pregunta cuatro repite la estructura de asociación simple con 8 incisos, con contenidos de recién nacido normal, enfermedades propias del recién nacido y el de alto riesgo. Las áreas examinadas fueron la semiología, clínica y epidemiología.

La pregunta cinco, de respuesta alternativa, incluye al igual que la uno tres temas que son las enfermedades renales con la infección urinaria, la glomerulonefritis y el síndrome nefrótico; las neurológicas con epilepsia, lesiones progresivas del sistema nerviosos central y síndrome convulsivo y por último el sistema hemolinfopoyético con la púrpura trombocitopénica idiopática, la de Sholein Henoch y el neuroblastoma

entre las enfermedades malignas. Las áreas exploradas fueron semiología, clínica, epidemiología y terapéutica.

La pregunta seis de selección múltiple exploró como área de la clínica los criterios diagnósticos de la Diabetes Mellitus.

La pregunta siete, de ensayo de respuestas cortas tipo completamiento, indagó con tres incisos las afecciones quirúrgicas como megacolon agangliónico, estenosis hipertrófica del píloro y etiología más frecuente del abdomen agudo en el recién nacido. Los tres ítems restantes fueron de enfermedades cardiovasculares como fiebre reumática, cardiopatías congénitas cianóticas y etiología de la hipertensión arterial en la adolescencia. Las áreas exploradas fueron semiología, clínica y básicas de la clínica.

Examen II.

La pregunta uno fue de ensayo de respuestas cortas tipo completamiento, en la que se explora el tema de recién nacido normal, de riesgo y el de enfermedades propias. El área de aprendizaje más examinada fue la clínica.

La pregunta dos examinó mediante respuesta alternativa contenidos de las infecciones respiratorias agudas altas, bajas y asma bronquial. Las áreas de aprendizaje fueron clínica, básicas de la clínica y terapéutica.

En la pregunta tres el diagnóstico diferencial, dentro del área de la clínica, entre la obstrucción intestinal alta y baja fue explorada con una pregunta de ensayo.

La pregunta cuatro fue combinada, tuvo ítems de respuesta alternativa, complemento simple y ensayo de respuestas cortas. Se comportó como de simulación progresiva porque se partió de una situación problemática de la cual se derivaron las diferentes preguntas y algunas de ellas sus respuestas dependían de la anterior. En ella se exploró contenidos de las enfermedades diarreicas agudas en el área de la semiología, clínica, básicas de la clínica y terapéutica. Fue una pregunta compleja para los estudiantes y profesores opinan fue bien trabajada.

La pregunta cinco examina tres temas con tres ítems cada uno. Los contenidos son las enfermedades cardiovasculares con las cardiopatías congénitas, las neurológicas

con lesiones progresivas, síndrome convulsivo e infección del sistema nervioso central y las renales con síndrome nefrótico, glomerulonefritis e infección del tractus urinario. Se exploraron las áreas de la clínica, básicas de la clínica y terapéutica.

La pregunta seis se comportó con igual estructura que la anterior, donde se comprobaron los temas generales como el crecimiento y desarrollo, la desnutrición proteico energética dentro del tema nutrición y la vacunación y sepsis en el tema de sepsis e inmunidad. Las áreas que se exploraron fueron la semiología, clínica, básicas de la clínica y la prevención.

La pregunta siete abarca dos temas que son enfermedades oncohematológica y endocrinas, el primero con contenidos de la púrpura de Sholein Henoch, la anemia ferropénica y la leucosis y el segundo con diabetes, hipotiroidismo y baja talla. El tipo de pregunta empleada fue la de complemento simple y se exploraron las áreas de semiología, epidemiología, básicas de la clínica y terapéutica.

II. Correspondencia entre los contenidos de la unidad curricular y su exploración en los exámenes.

El cuadro 1 muestra el tiempo dedicado según plan calendario de la asignatura a los contenidos temáticos y la frecuencia con que se exploran estos contenidos en los exámenes. Se observa que con la planeación de la evaluación en nuestra asignatura se logró una correspondencia entre estos dos parámetros, explorándose en ambos exámenes cada temática con un adecuado número de ítems. Esta correspondencia se corroboró cuando se aplicó el coeficiente de correlación de Spearman que para el examen I tuvo un valor de $r=0.84$ y para el II un $r=0,90$, ambos con una $p < 0,05$ IC 95 %.

Cuadro 1. Distribución del fondo de tiempo según programa y presencia de los incisos en las temáticas de los exámenes I y II.

Temáticas	Fondo de tiempo		Presencia en el examen I *		Presencia en el examen II **	
	Horas	%	Incisos	%	Incisos	%
Temas Generales	40	5.57	3	5.6	3	5.6
Nutrición	40	5.57	3	5.6	3	5.6
Inmunidad y Sepsis	40	5.57	3	5.6	3	5.6
Recién Nacido.	120	16.71	8	15.1	8	15.1
Enf. Digestivas/Desequilibrio hidromineral y ácido básico.	120	16.71	8	15.1	8	15.1
Enfermedades Respiratorias.	120	16.71	8	15.1	8	15.1
Enfermedades Endocrinas	40	5.57	5	9.4	3	5.6
Enf. Hemolinfopoyéticas	40	5.57	3	5.6	3	5.6
Enf. Cardiovasculares.	40	5.57	3	5.6	3	5.6
Enfermedades Renales	40	5.57	3	5.6	3	5.6
Enfermedades Nerviosas	40	5.57	3	5.6	3	5.6
Enfermedades Quirúrgicas	38	5.31	3	5.6	3	5.6
Total	718	100	53	100	51	100

Fuente: Datos del Autor.

* $r = 0,84$

** $r = 0,90$

Estos resultados no coinciden con los obtenidos en estudios de calidad realizados a exámenes de Morfofisiología II y Metodología de la Investigación y Estadística, en nuestra universidad, en los que no hubo correspondencia y existieron temáticas no exploradas o algunas de ellas con pocos ítems.^{42, 43.}

En la planeación de la evaluación lo primero es definir que debe ser evaluado y para ello hay que tener en cuenta las conductas, los problemas que los educandos deben ser capaces de enfrentar y las tareas en las que se espera que el educando sea competente.⁵

En nuestros exámenes se lograron explorar varios contenidos por temáticas y para su selección se tuvo en cuenta las entidades nosológicas más frecuentes en los servicios de salud y los problemas que el estudiante debe resolver de forma independiente. Así fue corroborado por el grupo de profesores y estudiantes entrevistados quienes coinciden que las afecciones y contenidos preguntados se encuentran dentro de los objetivos y en su mayoría son manejados en la práctica diaria.

III. Indicadores de calidad.

3.1. Análisis de la dificultad.

El índice de dificultad es la proporción de estudiantes que responden correctamente un inciso de un examen. Cuanto mayor sea esta proporción, menor será su dificultad. Lo que quiere decir que se trata de una relación inversa a mayor dificultad del ítem, menor será su índice.²⁹

El índice de dificultad para el examen I fue 0.81 clasificado como medianamente fácil sin embargo el II fue 0.74, lo cual nos infiere se trató de un examen de dificultad media. Este último resultado se acerca más a lo recomendado por Backhoff y es considerado positivo y esperado por nuestro grupo de trabajo luego de la planeación exhaustiva del instrumento.

Argudín Somonte y colaboradores en su trabajo "Índice de dificultad de Morfofisiología Humana I" obtuvo resultados similares a los encontrados en el examen I y expone que según el criterio de varios autores el nivel medio de dificultad de un examen debe oscilar entre el 50 y 60 %.⁸ La autora considera que el índice de dificultad del examen II se aproxima más a estas cifras propuestas y queda dentro del rango de valores de mayor calidad para este indicador.

En el anexo 4 y 5 se muestra el índice de dificultad por incisos, temáticas y preguntas de los exámenes I y II respectivamente, en los cuales nos basamos para la elaboración y discusión de los siguientes cuadros.

El cuadro 2 muestra el índice de dificultad por preguntas del examen I y II. En los exámenes anteriores por lo regular cada pregunta exploraba una temática, pero luego de la planeación existen preguntas que examinan hasta dos y tres temáticas

diferentes, así se comportaron las preguntas 1, 5 y 7 del examen I y las 5, 6 y 7 del examen II. Todos estos temas con un mismo fondo de tiempo en el plan calendario y con igual cantidad de incisos para cada una en los exámenes.

Cuadro 2. Grado de dificultad por preguntas de los exámenes.

Preguntas	Examen I		Examen II	
	Total de errores	Grado de dificultad	Total de errores	Grado de dificultad
Pregunta 1	282 (9)	0.84	422 (8)	0.74
Pregunta 2	224 (8)	0.85	441 (8)	0.73
Pregunta 3	203 (8)	0.88	152 (3)	0.69
Pregunta 4	250 (8)	0.84	321 (8)	0.76
Pregunta 5	401 (9)	0.77	392 (9)	0.78
Pregunta 6	133 (5)	0.86	524 (9)	0.71
Pregunta 7	475 (6)	0.59	352 (6)	0.71

Fuente: Datos del autor n= 192 est. Ítems: (53) n= 202 est. Ítems: (51)

En el examen I la mayoría de las preguntas alcanzaron un índice de dificultad medianamente fácil con excepción de la 5 y 7 que lograron un índice de 0.77 y 0.59 respectivamente comportándose como de dificultad media. Mientras que en el examen II la totalidad de las preguntas clasificaron de dificultad media.

Los resultados del examen I coinciden Con Ortiz Romero⁴² quien encontró en el análisis del instrumento escrito en la asignatura de Metodología de la Investigación y Estadística, preguntas tanto de dificultad media como medianamente fáciles. Sin embargo en nuestra búsqueda no aparecieron trabajos que arrojaran resultados como los hallados en el examen II.

A criterio de la autora en conjunto con los profesores la mayoría de las preguntas del examen I tienen un nivel de asimilación reproductivo, lo cual puede haber influido en este resultado, no así en el examen II que contiene mayor número de incisos de pensamiento y análisis médico.

Los profesores entrevistados opinan que para unir varias temáticas no bien delimitadas en una misma pregunta resulta complejo para la comprensión del

estudiante. Es cierto que los tipos de preguntas usadas en los exámenes para esta combinación fueron de complemento simple, de respuesta alternativa y de respuesta corta tipo completamiento, que permiten la exploración de varios contenidos diferentes. Pero la autora reflexiona que aún debemos mejorar la estructuración de estas preguntas pues cuando se escuchó a los estudiantes, refieren dificultades en la comprensión y elección de la respuesta correcta.

En la actualidad se aspira a que en la evaluación se ponga de manifiesto la integración de conocimientos que se adquieren en temas o capítulos separados de una misma disciplina lo que se conoce con el nombre de intradisciplinaridad. Para ello tenemos que perfeccionarnos más en la creación de preguntas que engloben varias temáticas.⁴⁴

Cuadro 3. Grado de dificultad por temáticas de los exámenes I y II.

Temáticas	Examen I		Examen II	
	Total de errores	Grado de dificultad	Total de errores	Grado de dificultad
Temas Generales	64 (3)	0.89	173 (3)	0.71
Nutrición	74 (3)	0.87	134 (3)	0.78
Sepsis e Inmunidad	144 (3)	0.75	217 (3)	0.64
Recién Nacido	250 (8)	0.84	422 (8)	0.74
Enfermedades Digestivas	203 (8)	0.88	321 (8)	0.76
Enf. Respiratorias.	224 (8)	0.85	441 (8)	0.73
Enfermedades Endocrinas	133 (5)	0.86	193 (3)	0.68
Enf. Hemolinfopoyéticas	58 (3)	0.90	159 (3)	0.74
Enf. Cardiovasculares.	224 (3)	0.61	116 (3)	0.81
Enfermedades Renales	134 (3)	0.77	97 (3)	0.84
Enfermedades Nerviosas.	209 (3)	0.64	179 (3)	0.70
Afecciones Quirúrgicas	251 (3)	0.56	152 (3)	0.69

Fuente: Datos del autor n= 192 est. Ítems: (53) n= 202 est. Ítems: (51)

En el cuadro 3 podemos apreciar el grado de dificultad por temáticas de los exámenes. En el examen I hay solo cinco de temáticas de dificultad media no así en el II en que 10 de las 12 temáticas tienen esta condición con excepción de los incisos

que exploraron las enfermedades renales y cardiovasculares que se comportaron medianamente fáciles con menor grado de dificultad.

Los temas de enfermedades quirúrgicas y cardiovasculares con 0.56 y 0.61 fueron los de mayor grado de dificultad presentes en el examen I. Ambos conformaron la pregunta siete. Para su elaboración se utilizaron ítems de ensayo de respuestas cortas tipo completamiento donde la autora detectó como deficiencia, la presencia de artículos que preceden los espacios en blanco, lo cual puede sugerir la palabra a completar y son indicadores de género y número.

A pesar de ello los resultados no fueron positivos. En el tema de afecciones quirúrgicas, a criterio del grupo focal de profesores, influyó que al confeccionar el ítem relacionado con el contenido de los vómitos en la estenosis hipertrófica del píloro no se especificó la característica que no debía faltar. Esta deficiencia en la elaboración del ítem provocó que pudiera existir más de una respuesta, sin embargo la clave no contaba con ellas. El análisis en el tema de enfermedades cardiovasculares arribó a la conclusión de que existió falta de preparación por parte de los estudiantes.

De este análisis se deriva la importancia que tiene la elección correcta de los tipos de ítems en correspondencia con el contenido que queremos explorar, la elaboración adecuada de los mismos así como la confección minuciosa y detallada de la clave de calificación. Lo antes referido nos llevará a una mayor calidad y éxito en la evaluación.

En el examen II, la sepsis e inmunidad es el tema de mayor grado de dificultad con 0.64. Este tema fue insertado en una pregunta problémica de la cual se derivaron incisos de selección múltiple tipo complemento simple en los que se abordaron contenidos de la vacunación, diagnóstico de la sepsis y su conducta a seguir.

Los ítems que exploraron la sepsis tuvieron un nivel de asimilación productivo y resultaron los de menor índice de dificultad de la temática. Los estudiantes alegaron se trató de una pregunta compleja, consideran el tema con insuficiente fondo de tiempo en el plan calendario y el de mayor dificultad para su dominio.

En el cuadro 4 mostramos la distribución de incisos según grado de dificultad esperado y real de los exámenes de acuerdo a la escala establecida por el criterio de los especialistas.

Cuadro 4. Distribución de incisos de los exámenes I y II según grado de dificultad esperado y real.

Incisos según grado de dificultad	Examen I				Examen II			
	Incisos esperados		Incisos reales		Incisos esperados		Incisos reales	
	No	%	No	%	No	%	No	%
Fáciles	3	5.6	20	37.7	3	5.9	2	3.9
Medianamente fáciles	10	18.9	14	26.4	10	19.6	13	25.5
De dificultad media	27	51.0	16	30.2	25	49.0	34	66.7
Medianamente difíciles	10	18.9	2	3.8	10	19.6	2	3.9
Difíciles	3	5.6	1	1.9	3	5.9	0	0
Totales	53	100	53	100	51	100	51	100

Fuente: Datos del autor.

Observamos que en el examen I no se cumplió con el pronóstico para la totalidad de los incisos. Se incrementaron en un mayor grado los fáciles y medianamente fáciles con 37.7 % de 5,6 % esperado y 26.4 % de 18.9 % respectivamente. No se logró alcanzar el porcentaje de incisos de dificultad media, medianamente difíciles y difíciles los cuales quedaron muy por debajo de lo esperado.

A diferencia del examen I en el II, apreciamos que los incisos medianamente fáciles y fáciles, se acercaron más al pronóstico. Los de dificultad media se elevaron en muy poca proporción de 49.0 % a 66.7 % aún aceptable. Los medianamente difíciles con 3.9 % de 19.6 % y los difíciles no presentes en el examen se quedaron muy por debajo de lo esperado.

Los resultados obtenidos evidencian que en el examen I no existió correspondencia entre los incisos esperados y los reales no así en el II en que se aprecia existe correspondencia entre estos dos indicadores. La aplicación del coeficiente de

correlación de Spearman, nos permitió corroborar lo antes expuesto, para el examen I con valor de $r = 0,158$ y para el II de $r = 0,738$ ambos con un IC 95 %.

Marín Hernández, en su investigación sobre la calidad del examen final de Morfofisiología Humana I en el estado Táchira, encontró total correspondencia entre los ítems esperados y reales a lo cual se acercan los resultados del examen II de nuestro estudio.³⁴

A pesar de estos resultados, la autora y profesores entrevistados, opinaron que aún se tienen que perfeccionar cada vez más las preguntas y tener en cuenta la necesidad de elaborar algunos incisos difíciles que puedan delimitar el verdadero estudiante de élite.

Los resultados obtenidos demuestran que solo con la planeación de la evaluación y la capacitación de nuestros profesores, en la elaboración correcta de preguntas e incisos se podrá lograr mejores instrumentos escritos en la asignatura. El análisis de la calidad de nuestros instrumentos permite resultados cada vez mejores y proporciona un mayor perfeccionamiento de la evaluación como componente indispensable del proceso enseñanza aprendizaje.

3.2. Análisis de la discriminación

El poder de discriminación, expresa la medida en que el ítem o pregunta discrimina entre los buenos y peores estudiantes. Se entiende que los buenos deben haber respondido en su mayoría correctamente el ítem y los estudiantes malos deben haberlo hecho incorrectamente.²⁸

En el Anexo 6 y 7 se muestran los índices de discriminación y coeficientes de discriminación por ítems y preguntas de los exámenes I y II respectivamente.

En el análisis del índice de discriminación, como se observa, en ambos existe un total predominio de los ítems con excelente poder discriminativo. Solo el inciso 5.7 del examen I que cuenta con un buen índice de discriminación con posibilidades de mejorarlo. De igual manera se comportaron las preguntas, en que solo la siete del examen I arroja igual posibilidad.

Estos resultados tan positivos llevó a la autora a determinar las razones discriminativas por incisos las cuales corroboraron que la totalidad de ellos, según su índice de discriminación, son aceptables, todos con valores superiores a uno. La autora concluyó que ambos instrumentos fueron capaces de diferenciar los estudiantes que tienen buenos resultados de aquellos con bajas calificaciones.

Autores como Backhoff²⁹ y Carrazana⁹ en sus investigaciones obtuvieron resultados similares en las que predominaron los ítems y preguntas excelentes y buenas. A diferencia del estudio realizado por González Sapsin⁴³ donde el predominio fue de ítems y preguntas regulares y pobres.

El coeficiente de discriminación se utiliza para saber si los alumnos de más alto rendimiento obtienen las respuestas correctas, es decir qué tanto valor predictivo tiene la pregunta y como se puede contribuir a las predicciones.

Al analizar el coeficiente de correlación biserial por cada inciso y pregunta del examen I (Anexo 6). Encontramos que de los 53 ítems 20 discriminan de forma regular y 11 lo hacen pobre con valores entre 0.15 – 0.28 y 0 – 0.14 respectivamente. Solo 15 tienen excelente poder discriminativo con valores superiores a 0.35. Con buena discriminación apenas 6 incisos con valores entre 0.26 – 0.35 y uno sin ninguna, con valores negativos.

En el examen II (Anexo 7) coincide que son 20 los incisos que discriminan de forma regular. Lo hacen de una forma buena 18 y solo 6 excelentes al igual que los ítems que discriminan de manera pobre. También encontramos un inciso con valores negativos.

En el cuadro 5 se relaciona el porcentaje de incisos de los exámenes según el coeficiente de correlación biserial. Se aprecia que en ambos exámenes el mayor número de ítems discriminaron de forma regular con 37.7 y 39.2 por cientos. En el examen II se encontró que la sumatoria de los incisos con excelente y buen poder discriminativo 47.1 % superan los de igual categoría del examen I 39.6 %. La

sumatoria de los que discriminaron de manera pobre y negativa predominaron en el examen I 22.7 % contra 13.7 %.

Cuadro 5. Porcentaje de ítems de los exámenes I y II según coeficiente de discriminación r_{pbis} .

Discriminación	Examen I		Examen II	
	No de ítems	% de ítems	No de ítems	% de ítems
Excelente	15	28.3	6	11.8
Buena	6	11.3	18	35.3
Regular	20	37.7	20	39.2
Pobre	11	20.8	6	11.8
Negativo	1	1.9	1	1.9
Total	53	100	51	100

Fuente: Datos del autor

Estos resultados permiten a la autora afirmar que el examen II cuenta con mayor poder predictivo y admite que los estudiantes de más alto rendimiento son los que han obtenido las más altas calificaciones.

Cuadro 6. Coeficiente de correlación del punto biserial (r_{pbis}) por preguntas.

Preguntas	Examen I		Examen II	
	R_{pbis}	Discriminación	R_{pbis}	Discriminación
Pregunta 1	0.20	Regular	0.47	Excelente
Pregunta 2	0.38	Excelente	0.23	Regular
Pregunta 3	0.15	Regular	0.40	Excelente
Pregunta 4	0.30	Buena	0.46	Excelente
Pregunta 5	0.24	Regular	0.42	Excelente
Pregunta 6	0.21	Regular	0.19	Regular
Pregunta 7	0.58	Excelente	0.55	Excelente

Fuente: Procesamiento Estadístico n=192

n=202

En el cuadro 6 se aprecia el comportamiento de las preguntas según su coeficiente de discriminación. En el examen I predominaron las preguntas que discriminaron de forma regular, que fueron la uno, tres, cinco y seis. Discriminaron de forma excelente la dos y la siete y la cuatro tuvo buen poder discriminativo. Mientras que en el examen II el predominio fue de las excelentes y solo la dos y seis tuvieron discriminación regular.

Autores como Carrazana y Backhoff consideran que el coeficiente de correlación punto biserial es una medida que combina la relación entre el criterio del reactivo y el nivel de dificultad.^{9, 29.}

En los exámenes analizados no se aprecia bien esta relación porque en el examen I no todas las preguntas con índices de dificultad medianamente fáciles tienen pobre o regular discriminación, como ocurre con la dos y la cuatro que tienen excelente y buena discriminación respectivamente. En el II no todas las preguntas discriminan de forma excelente a pesar de que la totalidad son de dificultad media. Lo hacen regular las preguntas dos y seis en las que se plantea la posibilidad que deben ser revisadas.

La autora considera que la variabilidad del coeficiente de discriminación de las preguntas puede deberse al tipo de clave que se utilizó para la decisión del patrón de aprobado desaprobado. Hay preguntas en las que se le asigna mayor valor a unos ítems que a otros y esto provoca que la calificación final de la pregunta no se corresponda con el número de ítems aprobados.

En la literatura revisada se considera que el coeficiente de discriminación es más sensible para detectar el comportamiento de los reactivos de un examen que el índice de discriminación.⁴⁵

En nuestra investigación se pone de ejemplo este criterio en los resultados del examen I, en el que la totalidad de las preguntas tuvieron un índice de discriminación entre excelente y bueno sin embargo el coeficiente en la mayoría de las preguntas fue regular.

En un estudio realizado en la Universidad Autónoma de México se obtuvo resultados similares pero a la inversa porque se logró un índice de discriminación pobre pero un buen coeficiente de discriminación.⁴⁵

3.3. Análisis de la confiabilidad

El coeficiente de confiabilidad alfa de Cronbach permite determinar la consistencia interna de un instrumento y brinda información relacionada con el aporte que hace cada pregunta a la consistencia general del examen.²⁸

Cuadro 7. Coeficiente Alfa de Cronbach por preguntas y exámenes.

Preguntas	Cronbach Examen I	Cronbach Examen II
Pregunta 1	0.71	0.63
Pregunta 2	0.68	0.69
Pregunta 3	0.71	0.67
Pregunta 4	0.67	0.63
Pregunta 5	0.68	0.65
Pregunta 6	0.73	0.67
Pregunta 7	0.66	0.64
Examen	0.72	0.69
Fuente: Procesamiento Estadístico. n=192		n=202

En el cuadro 7 se observan los valores del coeficiente alfa de Cronbach para cada una de las preguntas y de los exámenes como un todo. Los valores totales de cada instrumento fueron de 0.72 y 0.69 los que se acercan más a los estándares propuestos por el Colegio de médicos de Canadá donde consideran aceptables coeficientes entre 0.69 y 0.71. No así a los de Anastasi y a Weiner que aceptan como adecuados coeficientes que oscilan entre 0,80 y 0,85.⁴⁴

La Comisión Nacional de Evaluación de la Competencia y el Desempeño del Ministerio de Salud Pública proponen para este indicador valores superiores a 0.60 por lo que podemos concluir que los exámenes analizados tienen una confiabilidad aceptable.

En los resultados del coeficiente por preguntas se muestra que en el examen I al

Calidad de los exámenes certificativos de Pediatría.

eliminar la pregunta seis, el coeficiente alfa de Cronbach aumenta ligeramente respecto al total obtenido. Este resultado evidencia que dicha pregunta no aporta lo suficiente a la coherencia interna del examen y debe ser revisada y perfeccionada para su incorporación a próximos instrumentos escritos.

En el examen II la totalidad de las preguntas muestran valores iguales o inferiores al valor total del coeficiente alfa de Cronbach del examen, lo que nos permite concluir que todas las preguntas contribuyen de igual manera a la confiabilidad del instrumento.

A criterio de la autora y avalado por el grupo de profesores entrevistados esto se debe al grado de perfeccionamiento que se ha logrado en la elaboración de los instrumentos en la cátedra.

Calidad de los exámenes certificativos de Pediatría.

A. Resultados de la aplicación de los instrumentos y técnicas colectivas.

En ambos grupos focales los criterios de coincidencia fueron:

- Correspondencia entre los problemas de salud que se abordan en los exámenes y los problemas de salud a los que se enfrenta el médico general.
- Presencia de preguntas en los exámenes que exploran gran número de temáticas y contenidos sin delimitar su área lo que trae confusión al estudiante.
- La pregunta 3 del examen II, fue compleja para los estudiantes y evidenció falta de estudio.
- La mayoría de los estudiantes no lograron vencer el objetivo del diagnóstico y la conducta de la sepsis que se debió a la falta de estudio y pobre comprensión del tema.

B. Resultados del análisis de calidad de los instrumentos.

I. Características de los exámenes.

- Los dos exámenes están compuestos por 7 preguntas, el I consta de 53 incisos y el II de un total de 51. Ambos instrumentos exploraron la totalidad de las temáticas.

Calidad de los exámenes certificativos de Pediatría.

- Son exámenes mixtos con preguntas de desarrollo y de test objetivo con predominio de estas últimas, se incluyó una pregunta combinada en el examen II.
- Las áreas de aprendizaje que estuvieron representadas en ambos exámenes fueron la semiología, prevención, epidemiología, clínica, básicas de la clínica y terapéutica. El área más explorada fue la clínica en correspondencia con los objetivos del programa.

II. Correspondencia entre los contenidos de la unidad curricular y su exploración en los exámenes.

- Se exploraron en ambos exámenes la totalidad de los temas de la asignatura y existió una total correspondencia entre el fondo de tiempo dedicado a cada temática en el plan calendario y el número de incisos que exploraron cada temática en los exámenes, se corroboró a través del coeficiente de correlación de Spearman.

III. Indicadores de calidad.

3.1. Análisis de la dificultad.

- El examen I alcanzó un grado de dificultad medianamente fácil con un índice de dificultad de 0.81 mientras que el II logró ser de dificultad media con índice de 0.74.
- En el examen I las preguntas 1, 2, 3, 4 y 6 alcanzaron índices medianamente fáciles solo la 5 y 7 se comportaron con un grado de dificultad media. Mientras que en el II la totalidad de las preguntas clasificaron de dificultad media.
- De las 12 temáticas del programa analítico de Pediatría en el examen I solo hay cinco con índice de dificultad media mientras que en el II fueron 10 las que alcanzaron esta condición.
- En el examen I fueron las afecciones quirúrgicas y cardiovasculares las temáticas que tuvieron mayor grado de dificultad con 0.56 y 0.61 respectivamente. Fueron exploradas en una sola pregunta en la que se detectó deficiencias en su estructura y clave por el grupo focal y la autora.

- En el examen II, la temática de sepsis e inmunidad con 0.64 fue la más difícil, con un nivel de asimilación productivo y considerada por los estudiantes como una pregunta compleja y un tema de difícil dominio y escaso fondo de tiempo en el plan calendario.
- No existió correspondencia entre los incisos esperados y reales en el examen I a diferencia del II en que se logró correspondencia entre estos dos indicadores. Estos resultados fueron corroborados mediante el coeficiente de correlación de Spearman.

3.2. Análisis de la discriminación.

- En ambos exámenes existió predominio de los ítems y preguntas con excelente poder discriminativo, corroborado al hallar la razón discriminativa que permitió concluir que la totalidad de los incisos y preguntas fueron aceptables.
- Según el coeficiente de correlación biserial, en ambos exámenes, predominaron los ítems con discriminación regular con 37.7 y 39.2 por cientos.
- En el examen II la sumatoria de los incisos con excelente y buen poder discriminativo con 47.1 % superan los de igual categoría del examen I con 39.6 %. La sumatoria de los que discriminaron de manera pobre y negativa predominaron en el examen I con 22.7 % contra 13.7 %. La autora concluyó que el examen II tuvo mayor poder predictivo.
- Las preguntas con discriminación regular predominaron en el examen I, la 4 tuvo un buen poder discriminativo y la 2 y 7 excelente. Por el contrario, en el examen II, predominaron las preguntas con discriminación excelente y solo la 2 y la 6 lo hicieron de forma regular.
- En el examen I no se apreció bien la relación entre el índice de dificultad y el poder discriminativo porque no todas las preguntas con índices de dificultad medianamente fáciles tienen pobre o regular discriminación. La autora consideró que esta variabilidad puede deberse al tipo de clave que se utilizó para el patrón de aprobado desaprobado.

3.3. Análisis de la confiabilidad.

- Los valores del coeficiente de Cronbach para el examen I y II fueron de 0.72 y 0.69 respectivamente. Se concluyó que los instrumentos evaluativos analizados tienen una confiabilidad aceptable.
- En el análisis por preguntas se encontró que en el examen I la pregunta 6 no aportó lo suficiente a la coherencia interna del examen y debe ser revisada y perfeccionada. En el examen II todas las preguntas contribuyen de igual manera a la confiabilidad del instrumento.

CONCLUSIONES

1. Ambos exámenes analizados cuentan con siete preguntas y más de 50 ítems, se clasificaron como mixtos con predominio de las preguntas test objetivo, incluyeron la totalidad de las temáticas y la clínica fue el área de aprendizaje más explorada.
2. Se encontró total correspondencia entre los contenidos de la unidad curricular y su exploración en los dos exámenes.
3. El examen II tuvo mayor grado de dificultad con todas las preguntas de dificultad media, su temática más difícil fue la sepsis e inmunidad, explorada con un nivel de asimilación productiva y considerada compleja y de difícil dominio por los estudiantes.
4. Solo existió correspondencia entre los incisos esperados y reales en el examen II.
5. La totalidad de los ítems y preguntas de ambos exámenes fueron aceptables según índice de discriminación, sin embargo con el análisis del coeficiente de discriminación se comprobó que el examen II tuvo mayor poder predictivo.
6. Los exámenes analizados tienen una confiabilidad aceptable, solo la pregunta seis del examen I debe ser perfeccionada.

RECOMENDACIONES

1. Perfeccionar de forma sistemática el proceso de planeación de los instrumentos evaluativos en la cátedra de Pediatría que garantice una mayor calidad del proceso docente educativo.
2. Divulgar los resultados de la presente investigación en las diferentes formas de trabajo metodológico de la cátedra de Pediatría.

REFERENCIAS BIBLIOGRÁFICAS

1. Suarez de Lugo M. Conocimiento sobre evaluación de los aprendizajes de los evaluadores adscritos a las asignaturas del componente clínico pediátrico del programa de medicina. [Tesis para optar por el grado de master en educación en ciencias de la salud]. Venezuela: Universidad Centrooccidental Lisandro Alvarado, 2002.
2. Hernández Nodarse M. Perfeccionando los exámenes escritos: reflexiones y sugerencias metodológicas. OEI para la educación la ciencia y la cultura 2007; 41(4): 22-26
3. Liliana Ortiz M. La nueva concepción de la evaluación de aprendizajes en la educación de las Ciencias de la Salud. Rev Educ Cienc Salud. 2006; 3 (1): 6-13.
4. Salas Perea RS. La evaluación en la educación superior contemporánea. Biblioteca de Medicina. Volumen XXIV. UMSA, La Paz, Bolivia, 1998. [CD-ROM] Maestría en Educación Médica, ENSAP; 2005.
5. Salas Perea R. Planeamiento de un examen. Módulo Evaluación y recursos del aprendizaje Material de estudio de la Maestría en Educación Médica. En [CD-ROM] Maestría en Educación Médica, ENSAP: 2005.
6. Salas Perea R. Los instrumentos evaluativos. Módulo Evaluación y recursos del aprendizaje. Material de estudio de la Maestría en Educación Médica. En [CD-ROM] Maestría en Educación Médica, ENSAP: 2005.
7. Hatim Ricardo A. La educación de postgrado en el Sistema Nacional de Salud. Módulo de Postgrado. Material de estudio de la Maestría de Educación Médica, La Habana, 2002.
8. Argudín Somonte E, Díaz Rojas P, Leyva Sánchez E. Índice de Dificultad del examen de Morfofisiología Humana I. Rev Cubana Educ. Med Sup. 2011. Abr.-jun; 25(1)

9. Carrazana Lee A, Salas Perea RS, Ruiz Salvador AK. Nivel de dificultad y poder de Discriminación del examen diagnóstico de la asignatura Morfofisiología Humana I. *Educ Med Super* [Internet]; 2011 ene-mar [citado 12 mayo 2014]; 25(1): [Aprox. 12 p]. Disponible desde: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412011000100010&lng=es&nrm=iso&tlng=es
10. Salas Perea R. El proceso docente. En: *Proceso de Enseñanza-Aprendizaje. Material de estudio de la Maestría en Educación Médica*. La Habana: Escuela Nacional de Salud Pública; 2005.
11. Laudaluce Gutiérrez O. *Pedagogía*. La Habana: Editorial Ciencias Médicas; 2011.
12. Rivera Michelena, N. Enfoque psicopedagógico del proceso docente: Su impacto en el desarrollo de la personalidad. En: *Proceso de Enseñanza-Aprendizaje. Material de estudio de la Maestría en Educación Médica*. La Habana: Escuela Nacional de Salud Pública; 2005.
13. Salas Perea, RS. La identificación de las necesidades de aprendizaje en salud. En: *Educación de Postgrado. Material de estudio de la Maestría en Educación Médica*. La Habana: Escuela Nacional de Salud Pública; 2005.
14. Ahumada Acevedo P. *Principios y Procedimientos de Evaluación Educacional*. Chile: Universidad de Valparaíso; 1983.
15. Casanova M. El proceso de la evaluación educativa. *Ciberdocencia* [Internet]. 2003 Ago. [citado 7 febrero 2013]; [Aprox. 9 p]. Disponible en: <http://www.redesc.ilce.edu.mx/redescolar/cursosytalleres/evaluacion.htm>
16. De la Garza E. La Evaluación Educativa. *Rev Mex de Investigación Educativa* [Internet]; 2004 oct-dic [citado 15 febrero 2013]; 9(23): [Aprox. 11 p]. Disponible desde: <http://www.redalyc.org/src/inicio/ArtPdfRed.jsp?iCve=14002302>

17. Lafouercade PD. La calidad de la educación y el cuestionamiento de las pruebas estandarizadas de rendimiento en Estados Unidos de Norteamérica: las nuevas búsquedas. *Rev iberoam Educ* 1994; 2(3):53-61.
18. García Ramos L. Auto perfeccionamiento docente y creatividad. La Habana: Editorial Pueblo y Educación. 1996.
19. Cerda Gutiérrez H. La evaluación como experiencia total. Logros-objetivos-procesos. Competencia y desempeño. Santa Fe de Bogotá: Editorial Magisterio; 2000.
20. Carballo Carballo RM, Rico Romero L, Lupiáñez Gómez JL. Análisis de los ítems de las evaluaciones autonómicas de diagnóstico en España: 2008-2009. *Rev Iberoamericana de Educación Matemática* [Internet]; 2011 Junio [citado 2 de octubre de 2013]; (26): [Aprox. 16 p]. Disponible desde: http://www.fisem.org/web/union/index.php?option=com_content&view=article&d=32&Itemid=3
21. Ministerio de Salud Pública. Resolución Ministerial 15; 1988.
22. Cuba. Ministerio de Educación Superior. Reglamento de Trabajo Docente y Metodológico. Resolución 210; La Habana: Gaceta Oficial de la República de Cuba; 2007.
23. Salas Perea R. El examen escrito. Módulo Evaluación y recursos del aprendizaje. Material de estudio de la Maestría en Educación Médica. En [CD-ROM] Maestría en Educación Médica, ENSAP: 2005.
24. Díaz González NJ, García Carrasco F. Taller Examen Escrito. 2011 <http://www.conacem.org.mx/assets/11.pdf>
25. Pérez Espinosa L Martha. Comportamiento de los resultados del examen de competencia de Medicina General Integral. [Tesis para optar por el título de Master en Educación Médica Superior]. La Habana: ENSAP; 2003.

26. Salas Perea R. Calidad de los instrumentos de evaluación. Módulo Evaluación y recursos del aprendizaje. Material de estudio Maestría en Educación Médica. En: [CD-ROM] Maestría en Educación Médica. ENSAP: 2005.
27. Soler Huerta E, Sabido Sighler C, Sainz Vázquez L, Mendoza Sánchez H, Gil Alfaro I, González Solís R. Confiabilidad de un instrumento para evaluar aptitud clínica en residentes de medicina familiar. Archivos en Medicina Familiar 2005; 7(001): 14-17. [artículo en línea]. 2005 [citado 10 Mayo 2013]; [aprox. 5p.]. Disponible en: <http://www.medigraphic.com/pdfs/medfam/amf-2005/amf051e.pdf>
28. Díaz Rojas P, Leyva Sánchez E. Metodología para determinar la calidad de los instrumentos de evaluación. Educ Med Super. [Internet] 2013 [citado 18 Mayo 2013]; 27(2): [aprox. 18 p]: Disponible en: http://scielo.sld.cu/scielo.php?script=sci_issues&pid=0864-2141&lng=es&nrm=iso.
29. Backhoff E., Larrazolo N., Rosas M. Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). Revista Electrónica de Investigación Educativa. 2000. En [CD-ROM] Maestría Educación Médica. ENSAP; 2005.
30. Pérez Espinosa L Martha. Comportamiento de los resultados del examen de competencia de Medicina General Integral. [Tesis para optar por el título de Master en Educación Médica Superior]. La Habana: ENSAP; 2003.
31. Diamant AE. Propuesta para un análisis comparativo de instrumentos de evaluación de logros académicos. Ministerio de Cultura y Educación República Argentina [Citado el 20 enero de 2012] Disponible en: <http://www.ince.mec.es/cumbre/d1/04.htm>
32. Propuesta de indicadores de calidad para la evaluación de la calidad de las preguntas en los exámenes de opción múltiple. Presentación: Latin Educa 2004.com. Primer congreso virtual latinoamericano De Educación a distancia.

Del 23 de marzo al 4 de abril del 2004. [Citado 2 febrero 2012]. Disponible en:http://www.ateneonline.net/datos/91_03%20PONENCIA%20CORICAHERNANDEZ-HOLLOWAY-DIMOU.pdf.

33. Moreno Montañez M. Caracterización del examen estatal escrito de la especialidad en Medicina General Integral. Primera convocatoria. Venezuela. [Tesis para optar por el título de Master en Educación Médica Superior] La Habana: ENSAP; 2007.
34. Marín Hernández D. Estructura y valoración de la calidad del examen final de Morfofisiología Humana I en el Estado Táchira. Curso 2005 – 2006. [Tesis para optar por el título de Master en Educación Médica Superior] La Habana: ENSAP; 2007.
35. Belle J, Handfield-Jones R, Rainsberry P, Brailovsky CA. Certification Examination of the College of Family Physician of Canada. Part 4 Simulated Office Orals. *Can Fam Physician* 1996; 42(8):1539-1548.
36. Handfield-Jones R, Belle J, Rainsberry P, Brailovsky CA. Certification Examination of the College of Family Physician of Canada. Part 3. Short-answer management problems. *Can Fam Physician* 1996; 42(7):1353-1361.
37. Weiner EA, Stewart BJ. *Assessing individuals*. Little Brown, Boston. 1984.
38. Van Der Vleuten CPM, Swanson DB. Assessment of Clinical Skills with standardized patients: state of the art. *Teach Learn Med*. 1990; 2:58-76.
39. Charlin B, Leduc C, Brailovsky CA, Blonin D, Hwon R. The diagnostic schema questionnaire: A new tool to assess a specific dimension of clinical competence. *Advances in Medical Education*. 1997: 354-356.
40. Tamblyn R, Abrahamowicz M, Brailovsky CA, Grand'Mason P. Association between Licensing Examination Scores and Resource use and Quality of Care in Primary Care Practice. *JAMA*. 1998; 280 (11): 989 – 996.

41. López Espinoza GJ, Quintana Mugica R, Rodríguez Cruz O, et al. El profesor principal y su preparación para diseñar instrumentos de evaluación escritos. EDUMECENTRO [Internet]. 2014 [citado 6 May 2014];6(2):[aprox. 8p.]. Disponible en: <http://www.revedumecentro.sld.cu/index.php/edumc/article/view/354/html>
42. Ortiz Romero G. Calidad del examen final de la asignatura Metodología de la Investigación y Estadística. Carrera de Medicina. Curso 2010-2011. [Tesis para optar por el título de Master en Educación Médica Superior] Holguín; 2012.
43. González Sapsin K. Calidad del examen final de la asignatura Morfofisiología Humana II. Curso 2009-2010. [Tesis para optar por el título de Master en Educación Médica Superior] Holguín; 2012.
44. Vicedo Tomey A. La integración de conocimientos en la educación médica. Educ Med Super v.23 n.4 Ciudad de la Habana oct.-dic. 2009.
45. Sánchez González G, Tapia Pérez GG. Descripción del nivel de facilidad y poder de discriminación del examen de inferencia estadística en métodos estadísticos en medicina veterinaria y zootecnia. Universidad Autónoma de México. Facultad de Medicina Veterinaria y Zootecnia. [en línea] [Fecha de acceso 2 de abril de 2014] Disponible en: http://www.facmed.unam.mx/sem/jem2010/Abstracts/trabajos_orales.pdf

ANEXO 1.

Consulta a Especialistas.

Estimados profesores:

La presente consulta se realiza como parte del desarrollo de una investigación en el campo de la evaluación.

Necesitamos de ustedes su sincera opinión en relación con las preguntas que les realizamos. Su opinión será importante para mejorar el trabajo docente metodológico de los profesores que imparten la asignatura Pediatría y contribuir de esta manera al perfeccionamiento del programa de estudio en función de elevar los niveles de competencia y desempeño de los futuros egresados. Sus opiniones solo serán usadas con fines investigativos.

Guía para la Discusión:

1. ¿Qué área de aprendizaje están exploradas en cada uno de los incisos del examen?
2. ¿Qué tipo de items fue el utilizado en la confección de cada inciso del examen?
3. ¿Consideran que el formato de las preguntas empleadas está acorde a éste tipo de examen? ¿Consideran que existió una adecuada proporción de los tipos de preguntas?
4. ¿Qué contenidos temáticos de la asignatura y qué objetivos fueron explorados en cada inciso del examen? ¿Consideran que se le da salida a los objetivos generales del médico general básico?
5. ¿Cómo clasificarían los incisos y preguntas del examen de Pediatría de acuerdo a: difíciles, medianamente difíciles, dificultad media, medianamente fáciles y fáciles, atendiendo al porcentaje de estudiantes que aprobaron los incisos y preguntas?

ANEXO 2.

Grupo Focal para Profesores.

Objetivo: Obtener información con relación a la calidad de los exámenes ordinarios de Pediatría del curso 2013 - 2014.

Estimados profesores:

La presente entrevista se realiza como parte del desarrollo de una investigación en el campo de la evaluación.

Necesitamos de ustedes su sincera opinión en relación con las preguntas que les realizamos. Su opinión será importante para mejorar el trabajo docente metodológico de los profesores que imparten la asignatura Pediatría y contribuir de esta manera al perfeccionamiento del programa de estudio en función de elevar los niveles de competencia y desempeño de los futuros egresados. Sus opiniones solo serán usadas con fines investigativos.

Guía para la Discusión:

1. ¿Los problemas de salud que se abordan en el examen están en correspondencia con los problemas de salud concretos de la práctica del futuro profesional?
2. ¿La ponderación otorgada en la clave de calificación a las preguntas consideran es la más adecuada?
3. Valoración de los resultados obtenidos de cada una de las preguntas del examen y sus consideraciones de las causas de estos resultados.
4. Consideran ustedes que existe adecuada concordancia entre las preguntas del instrumento y el contenido que se evalúa.

ANEXO 3.

Grupo Focal de Estudiantes.

Objetivo: Obtener información con relación a la calidad de los exámenes ordinarios de Pediatría del curso 2013 - 2014.

Estimados estudiantes:

La presente entrevista se realiza como parte del desarrollo de una investigación en el campo de la evaluación.

Necesitamos de ustedes su sincera opinión en relación con las preguntas que les realizamos. Su opinión será importante para mejorar el trabajo docente metodológico de los profesores que imparten la asignatura Pediatría y contribuir de esta manera al perfeccionamiento del programa de estudio en función de elevar los niveles de competencia y desempeño de ustedes como futuros egresados. Sus opiniones solo serán usadas con fines investigativos.

Guía para la Discusión:

1. ¿Los contenidos recibidos durante las diferentes rotaciones les permitieron responder los diferentes incisos que conformaron el examen?
2. ¿Consideran que el tipo de preguntas empleadas se corresponden con la forma en que recibieron y le evaluaron a ustedes los contenidos?
3. ¿Creen ustedes que los temas de la asignatura ofrecidos y evaluados en el examen le permiten su desempeño como médico general?
4. Principales causas de los resultados obtenidos en cada una de las preguntas del examen.

Anexo 4. Grado de dificultad por incisos, temáticas y preguntas del examen I.

5-7	Enfermedades	0.94		
5-8	Hemolinfopoyéticas	0.93	0.90	Nivel de dificultad. pregunta
5-9	Temáticas	0.82	0.90	
incisos		incisos	temática	
6-a	Enfermedades	0.92	0.89	0.86
6-b	Terapias	0.95		
6-c	Enfermedades	0.96		
6-d	Terapias	0.97		
7-1	Nutrición	0.88	0.87	0.84
7-2	Afecciones quirúrgicas	0.82	0.56	0.59
7-3		0.40		
7-4	Señales de alarma	0.88	0.75	
7-5	Cardiovasculares	0.93	0.61	
2-1		0.83		
2-2	Sistema Respiratorio	0.94	0.85	0.85
2-3		0.75		
2-4		0.91		
2-5		0.75		
2-6		0.94		
2-7		0.81		
2-8		0.91		
3-1	Enfermedades Digestivas	0.84	0.88	0.88
3-2		0.95		
3-3		0.98		
3-4		0.71		
3-5		0.85		
3-6		0.95		
3-7		0.90		
3-8		0.83		
4-a	Recién Nacido	0.93	0.84	0.84
4-b		0.64		
4-c		0.81		
4-d		0.94		
4-e		0.70		
4-f		0.85		
4-g		0.93		
4-h		0.89		
5-1	Enfermedades Renales	0.98	0.77	0.77
5-2		0.58		
5-3		0.74		
5-4	Enfermedades Nerviosas	0.52	0.64	
5-5		0.56		
5-6		0.83		

Fuente: Datos del autor

Anexo 5. Grado de dificultad por incisos, temáticas y preguntas del examen II.

Preg. e incisos	Temáticas	Nivel de dificult. incisos	Nivel de dificult. temática	Nivel de dificult. pregunta
1-a	Recién Nacido	0.79	0.74	0.74
1-b		0.70		
1-c		0.66		
1-d		0.66		
1-e		0.78		
1-f		0.66		
1-g		0.86		

Calidad de los exámenes certificativos de Pediatría.

1-h		0.80		
2-a	Sistema Respiratorio	0.77	0.73	0.73
2-b		0.47		
2-c		0.72		
2-d		0.72		
2-e		0.74		
2-f		0.82		
2-g		0.79		
2-h		0.79		
3-a		Afecciones quirúrgicas		
3-b	0.49			
3-c	0.79			
4-a	Enfermedades Digestivas	0.81	0.76	0.76
4-b		0.56		
4-c		0.85		
4-d		0.86		
4-e		0.80		
4-f		0.77		
4-g		0.57		
4-h		0.89		
5-1	Enfermedades Cardiovasculares	0.66	0.81	0.78
5-2		0.91		
5-3		0.85		
5-4	Enfermedades Nerviosas	0.72	0.70	
5-5		0.65		
5-6		0.74		
5-7	Enfermedades Renales	0.89	0.84	
5-8		0.70		
5-9		0.93		
6-1	Temas Generales	0.89	0.71	0.71
6-2		0.57		
6-3		0.69		
6-4	Nutrición	0.59	0.78	
6-5		0.88		
6-6		0.87		
6-7	Sepsis e Inmunidad	0.74	0.64	
6-8		0.53		
6-9		0.66		
7-1	Enfermedades Hemolinfopoyéticas	0.81	0.74	0.71
7-2		0.83		
7-3		0.57		
7-4	Enfermedades Endocrinas	0.75	0.68	
7-5		0.73		
7-6		0.57		

Fuente: Datos del autor.

Anexo 6. Índice y Coeficiente de discriminación por incisos y preguntas del examen I.

Preg. e incisos	Temáticas	Índice de discriminación		Coeficiente de discriminación	
		Incisos	Preguntas	Incisos	Preguntas
1-a	Temas Generales	0.48	0.52	0.14	0.20
1-b		0.53		0.03	
1-c		0.69		0.25	
1-d	Nutrición	0.45		0.08	
1-e		0.55		0.11	
1-f		0.64		0.28	

Calidad de los exámenes certificativos de Pediatría.

1-g	Sepsis e Inmunidad	0.77		0.36		
1-h		0.57		0.19		
1-i		0.47		0.15		
2-1	Sistema Respiratorio	0.57	0.78	0.33	0.38	
2-2		0.54		0.18		
2-3		0.69		0.45		
2-4		0.54		0.21		
2-5		0.74		0.45		
2-6		0.48		0.19		
2-7		0.59		0.40		
2-8		0.47		0.21		
3-1	Enfermedades Digestivas	0.50	0.78	0.22	0.15	
3-2		0.51		0.16		
3-3		0.45		0.02		
3-4		0.54		0.07		
3-5		0.51		0.16		
3-6		0.48		0.22		
3-7		0.60		0.13		
3-8		0.58		0.05		
4-a	Recién Nacido	0.44	0.66	0.08	0.30	
4-b		0.79		0.42		
4-c		0.53		0.23		
4-d		0.48		0.18		
4-e		0.73		0.47		
4-f		0.75		0.36		
4-g		0.47		0.18		
4-h		0.62		0.37		
5-1	Enfermedades Renales	0.48	0.57	0.06	0.24	
5-2		0.65		0.36		
5-3	0.66	0.45				
5-4	Enfermedades Nerviosas	0.81		0.29		
5-5		0.52		0.15		
5-6		0.66		0.22		
5-7		Enfermedades Hemolinfopoyéticas		0.37		-0.20
5-8				0.54		0.15
5-9	0.60	0.34				
6-a	Enfermedades Endocrinas	0.59	0.81	0.29	0.21	
6-b		0.51		0.20		
6-c		0.69		0.27		
6-d		0.44		0.12		

Calidad de los exámenes certificativos de Pediatría.

6-e		0.50		0.16	
7-1	Afecciones quirúrgicas	0.63	0.33	0.37	0.58
7-2		0.82		0.44	
7-3		0.95		0.15	
7-4	Enfermedades Cardiovasculares	0.96		0.37	
7-5		0.64		0.46	
7-6		0.76		0.37	

Fuentes: Datos del autor.

Anexo 7. Índice y Coeficiente de discriminación por incisos y preguntas del examen II.

Preg. e incisos	Temáticas	Índice de discriminación		Coeficiente de discriminación	
		Incisos	Preguntas	Incisos	Preguntas
1-a	Recién Nacido	0.71	0.70	0.27	0.47
1-b		0.83		0.23	
1-c		0.86		0.31	
1-d		0.67		0.23	
1-e		0.61		0.20	

Calidad de los exámenes certificativos de Pediatría.

1-f		0.91		0.47	
1-g		0.74		0.29	
1-h		0.83		0.33	
2-a	Sistema Respiratorio	0.71	0.76	0.07	0.23
2-b		0.83		0.16	
2-c		0.64		0.22	
2-d		0.73		0.07	
2-e		0.78		0.29	
2-f		0.78		0.19	
2-g		0.75		0.13	
2-h		0.76		0.27	
3-a	Afecciones quirúrgicas	0.74	0.72	0.33	0.40
3-b		0.86		0.33	
3-c		0.78		0.40	
4-a	Enfermedades Digestivas	0.74	0.72	0.26	0.46
4-b		0.78		0.16	
4-c		0.74		0.41	
4-d		0.74		0.28	
4-e		0.86		0.26	
4-f		0.90		0.49	
4-g		0.92		0.41	
4-h		0.59		-0.05	
5-1	Enfermedades Cardiovasculares	0.85	0.66	0.23	0.42
5-2		0.74		0.18	
5-3		0.68		0.12	
5-4	Enfermedades Nerviosas	0.93		0.29	
5-5		0.83		0.39	
5-6		0.75		0.22	
5-7	Enfermedades Renales	0.69		0.20	
5-8		0.72		0.23	
5-9		0.68		0.17	
6-1	Temas Generales	0.66	0.72	0.21	0.19
6-2		0.57		0.16	
6-3		0.77		0.23	
6-4	Nutrición	0.78		0.24	
6-5		0.63		0.05	
6-6		0.65		0.05	
6-7	Sepsis e Inmunidad	0.70		0.22	
6-8		0.77		0.16	
6-9		0.79		0.31	

Calidad de los exámenes certificativos de Pediatría.

7-1	Enfermedades Hemolinfopoyéticas	0.77	0.62	0.32	0.55
7-2		0.74		0.25	
7-3		0.84		0.32	
7-4	Enfermedades Endocrinas	0.84		0.28	
7-5		0.86		0.29	
7-6		0.94		0.32	

Fuente: Datos del autor.