

Numeralia de los EGEL

A continuación se presenta, en forma gráfica, la información sobre el número de sustentantes de los EGEL de 1994 a 2017 (Figura 2), así como el número de instituciones que aplicaron algún EGEL en el mismo período (Figura 3).

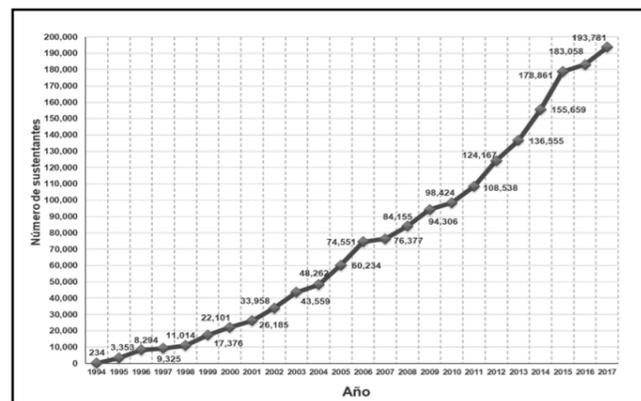


Figura 2. Número de sustentantes de los EGEL 1994-2017

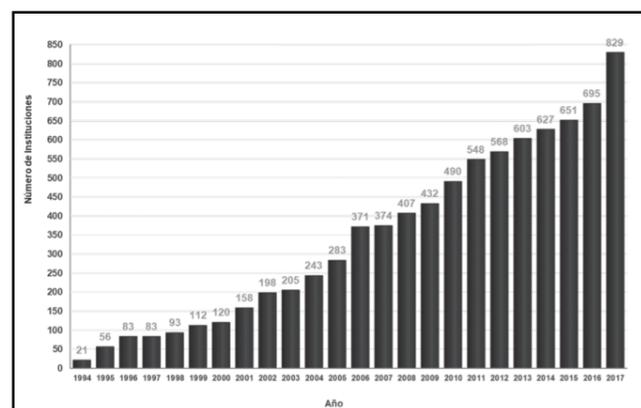


Figura 3. Número de Instituciones que aplican los EGEL 1994-2017

Los Exámenes Diagnósticos de la Licenciatura (EXDIAL)

Adicionalmente a los EGEL, en el Ceneval también se elaboran los Exámenes Diagnósticos e Intermedios para la Licenciatura. Los primeros son instrumentos de evaluación de cobertura nacional que permiten identificar si los estudiantes de la licenciatura cuentan con los conocimientos, habilidades y competencias comunes a un conjunto de licenciaturas o incluso, a todas ellas y que son indispensables para su adecuada formación profesional. Un examen representativo de lo anterior es el examen de competencia comunicativa y pensamiento crítico (ECC y PEC). Un examen intermedio es el EXIL-CBI, el cual es un

examen diagnóstico de cobertura nacional que tiene como propósito identificar el nivel de dominio que poseen los futuros ingenieros en la fase intermedia de su licenciatura, respecto a los conocimientos y habilidades intelectuales considerados en las ciencias básicas de Ingeniería, Matemáticas, Física y Química.

Cuestionario de Contexto de los EGEL

Cada estudiante que contesta un EGEL debe también responder un cuestionario de contexto. El objetivo principal de este cuestionario es identificar los perfiles de las escuelas según las opiniones que tienen sus egresados de éstas, en cuanto a la infraestructura, calidad docente, plan de estudios y otras características de las instituciones.

Los retos relacionados con este instrumento son:

- Perfeccionar el sustento conceptual de cada una de las variables incorporadas en los cuestionarios, integrando hallazgos recientes que contribuyan a alcanzar los objetivos de cada instrumento.
- Instrumentar estrategias de difusión sobre la importancia de los cuestionarios de contexto entre las instituciones usuarias, a fin de que la información generada les pueda ser útil en su proceso de toma de decisiones, así como en el desarrollo de programas y políticas públicas.

Nueva generación de los EGEL

Por último, cabe mencionar que actualmente se ha comenzado el diseño y conceptualización de los EGEL de tercera generación, los cuales deberán ser piloteados en 2019.

BIBLIOGRAFIA

1. Ceneval. (1998). *Acerca del Ceneval y los Exámenes Generales para el Egreso de la Licenciatura EGEL*. México: Ceneval.
2. Ceneval (2012). *Reglamento de los Consejos Técnicos de los exámenes*. México: Ceneval.
3. Hernández, J. y Delgado, L. (2009). *Encuesta nacional para la validación social de los perfiles profesionales de los Exámenes Generales para el Egreso de la Licenciatura (EGEL)*. Marcos de referencia 4. México: Ceneval
4. Hernández Uralde, J. (2013). *Exámenes Generales para el Egreso de la Licenciatura*. México: Ceneval.
5. Raymond, M. R. y Neustel, S. (2006). Determining the content of credentialing examinations. En S.M. Downing & T.M.M. Haladyna, *Handbook of test development* (pp. 181-223). Mahwah, NJ: Lawrence Erlbaum Associates.

Artículo Original

Habilitación Profesional: Caso Ecuador

Ana Lucía Ruano, Alejandro Vizuite, Juan Carlos Moreno, Willian Quispe Revelo.

Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior, Ceaaces- ECUADOR.

DIRECCIÓN PARA CORRESPONDENCIA: Dra. Ana Lucía Ruano, Ceaaces, German Alemán y Javier Aráuz E11-32, Quito 170135, ana.ruano@ceaaces.gob.ec.

RESUMEN

Históricamente en los diversos sistemas educativos, la principal herramienta para evaluar los resultados de aprendizaje corresponde a pruebas, test o exámenes, que pueden ser de orden escrito, verbal o demostraciones prácticas. Estos instrumentos son utilizados a diferentes escalas, desde el aula de clase hasta los exámenes a nivel local, nacional o regional y generalmente sirven de base para inferir sobre los conocimientos y aptitudes de los evaluados. Esta clase de exámenes se encuentran sujetos a ciertas dificultades y limitaciones, como el determinar a priori, el nivel de aprendizaje - habilidad- de la cohorte o grupo de personas a ser evaluadas. Una de las actividades con mayor dificultad en la gestión de la aplicación de estos exámenes es determinar el punto de corte que establece si una persona adquirió, o no, las habilidades y competencias para ejercer la profesión, para esto se utilizan varias metodologías (standard setting). En este artículo se detalla y analizan los métodos utilizados por el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior-Ceaaces, para determinar estándares en los exámenes que se aplican como parte de la certificación para la habilitación profesional en el Ecuador.

Palabras clave: Exámenes, estándares, habilitación profesional.

ABSTRACT

Historically in the different educational systems, the main tool to evaluate the learning results corresponds to tests or exams, which can be written, verbal or practical demonstrations. These instruments are used at different scales, from the classroom to the exams at the local, national or regional level and generally serve as a basis for inferring knowledge and skills of the evaluated. This type of exams are subject to certain difficulties and limitations, such as determining a priori, the level of learning -ability- of the cohort or group of people to be evaluated. One of the activities with greater difficulty in the management of the application of these examinations is to determine the cut-off point that establishes whether a person acquired, or not, the skills and competencies to practice the profession, for this several methodologies are used (standard setting). This article details and analyzes the methods used by the Council of Evaluation, Accreditation and Quality Assurance of Higher, to determine standards in the exams that are applied as part of the certification for professional qualification in Ecuador.

Key words: Tests, standards, professional qualification.

INTRODUCCIÓN

Los exámenes de gran escala y de alto impacto han sido utilizados históricamente a nivel mundial, nadie puede negar que han contribuido al desarrollo de la evaluación educativa, sin embargo como en muchas metodologías de evaluación, existen criterios diversos que analizan su valor real y sus potenciales efectos negativos.¹

El análisis de las competencias en los diferentes instrumentos utilizados para la evaluación parte de su conceptualización, así pues, un examen de gran escala se desarrolla en nivel local, nacional o regional, considerando un gran número de evaluados. Un examen de alto impacto es aquel cuyo resultado es empleado para tomar decisiones importantes que conllevan consecuencias positivas o negativas en quienes los sustentan.²

A diferencia de los exámenes a gran escala, los de alto impacto pueden o no ser administrados a grandes poblaciones. Por ejemplo, en el contexto educativo, los resultados de los exámenes de aptitud, certificación y dominio son frecuentemente empleados por las partes interesadas como criterios de egreso/ingreso tanto en la educación media como superior. Un examen de alto impacto debe tener una validez indiscutible, esta característica es indispensable para tomar decisiones equitativas y justas basadas en la información veraz de los resultados respecto a los conocimientos, habilidades y competencias demostradas por los evaluados.³

En el Ecuador, el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior-Ceaaces, es el ente encargado, por ley, de garantizar que las universidades y sus carreras, así como los institutos técnicos y tecnológicos, articulen procedimientos que permitan formar profesionales que atiendan y contribuyan de manera adecuada a la sociedad, a través de procesos de evaluación de entornos y resultados de aprendizaje.

Uno de los principales ejes que maneja el consejo es el aseguramiento de la calidad de la educación superior en Ecuador, que lo realiza a través de varias aristas y con la participación de la comunidad académica. Dentro de sus competencias el Ceaaces, evalúa a las carreras o programas de las Instituciones de Educación Superior (IES), a través de dos ámbitos; el primero evalúa los procesos y la gestión que se realizan para garantizar las condiciones mínimas en las que se desarrollan los procesos de enseñanza aprendizaje y el segundo está enfocado en evaluar los resultados del aprendizaje, a través de una prueba denominada Examen Nacional de Evaluación de Carreras (ENEC). Como parte del aseguramiento de la calidad propuesto, es necesario garantizar que los estudiantes y/o profesionales de las carreras denominadas de interés público, tengan los conocimientos mínimos para ejercer la profesión, esto se lo realiza a través de una prueba o test denominado Examen de Habilitación para el Ejercicio Profesional (EHEP).

Los exámenes que aplica el Ceaaces (EHEP y ENEC), son pruebas de opción múltiple aplicadas a los estudiantes de último año de formación y en el caso del EHEP, permite también evaluar a profesionales que obtuvieron su título de grado en el extranjero y que deseen ejercer la profesión en nuestro país.

El EHEP y el ENEC, se han constituido en exámenes de alto impacto, puesto que constituyen uno de los requisitos para la habilitación profesional en las carreras de interés público según nuestra legislación. Estos exámenes mantienen una estructura que se fundamenta en la realidad ecuatoriana y que ha sido analizada por las IES que ofertan carreras en el área de salud evaluadas y los demás actores del sistema de educación superior, los que determinaron cuatro ejes obligatorios en todos los componentes del examen que son:

- El perfil del médico ecuatoriano.
- Las mallas curriculares.
- Morbi-mortalidad.
- Modelo de Atención Integral de Salud (MAIS)

El examen constituido con todos estos elementos, se diferencia de los exámenes utilizados con fines de certificación profesional en la región.

Una vez aplicada la evaluación, los diferentes escenarios a los que se deberá enfrentar un sustentante de estos exámenes a gran escala deben estar fundamentados en garantizar la idoneidad del instrumento de evaluación, así como en un análisis detallado de los resultados plasmados en la calificación final obtenida por cada evaluado y por toda la cohorte.

Uno de los elementos importantes para garantizar la objetividad de los resultados es determinar el estándar o punto de corte con el que un sustentante aprueba o no el examen.

Para comprender de mejor manera los diferentes métodos escogidos y utilizados es importante responder las siguientes preguntas:

Qué es un estándar?

Un estándar es un nivel deseado de desempeño, puede ser definido de manera conceptual o empírica, son marcados por expertos y si hablamos de exámenes y su análisis puede obtenerse a nivel individual o grupal. Estrictamente hablando, no existen métodos para el establecimiento de estándares, los que existen son métodos para el establecimiento de los puntos de corte.⁴

Qué es un punto de corte?

Son los valores numéricos que operacionalizan los niveles de desempeño establecidos. Existen diversos métodos orientados a esta tarea. Cada método posee además un sistema preestablecido y racional de reglas destinadas a identificar cuál es el valor numérico adecuado que separa un nivel de desempeño de otro. Algunos de los más conocidos son el método de

Angoff (1971) y el método Bookmark (Lewis, DM. 1996-1999).⁵⁻⁷

Cualquier método utilizado para establecer los puntos de corte y el establecimiento de estándares, tendrá que contener al menos estos pasos:⁸

- Selección del método.
- Selección de jueces o especialistas según el área del conocimiento valorada.
- Definición del individuo limítrofe, tomando en cuenta el desempeño de toda la cohorte evaluada.
- Capacitación de los jueces o especialistas en el método seleccionado.
- Obtención de los puntos de corte establecidos por los grupos de jueces o especialistas.
- Retroalimentación de puntos de corte preliminares.

De entre los varios métodos existentes para establecer el estándar deseado, la elección de uno u otro dependerá de:

- Heterogeneidad de las preguntas utilizadas en la evaluación (preguntas de opción múltiple suelen ser evaluados con el método Angoff, mientras que cuando se utiliza una combinación de ítems de opción múltiple y de preguntas abiertas se suele recomendar el método Bookmark)
- Tiempo disponible para realizar las evaluaciones
- Experiencia previa con el método (la utilización previa del método facilita el trabajo de campo posterior y reduce la cantidad de tiempo a las tareas, por lo tanto, a veces es más ventajoso).
- Evidencias de validez del método (algunos investigadores han cuestionado la validez de métodos como Angoff o método de grupos contrastados).⁹

Las metodologías utilizadas para determinar el estándar en cada aplicación de los exámenes desarrollados por el Ceaaces son descritas y analizadas en el presente documento con el fin de que sean conocidas por los actores involucrados en estos procesos.

METODOLOGÍA

El presente trabajo tiene como objetivo dar a conocer y comparar las metodologías empleadas por el Ceaaces para determinar los estándares en los exámenes aplicados, específicamente en el área de la salud, denominado Examen de habilitación para el ejercicio profesional (EHEP), el que está orientado a que los sustentantes adquieran la certificación que forma parte de los requisitos para obtener la

habilitación para el ejercicio profesional en el Ecuador.

Se describirá como caso práctico, el desarrollo de las metodologías aplicadas para determinar el punto de corte, con datos reales de las IES involucradas y de sus sustentantes. Se analizó una aplicación realizada en el año 2018, de una carrera del área de la salud. Se valoró las metodologías con 1317 sustentantes provenientes de una carrera presente a nivel nacional.

Las metodologías utilizadas se describen a continuación:

Metodologías Absolutas

Estas metodologías utilizan el criterio de jueces o expertos en el área del conocimiento que se evalúa en el examen, los que son capacitados con la metodología antes de iniciar el análisis en el taller de trabajo. Estos métodos utilizados son:

Bookmark Standard Setting

De los métodos aplicados para garantizar que el examinado domine los contenidos mínimos establecidos previos al examen, uno de los más utilizados para definir puntos de corte ha sido el que se conoce como procedimiento de definición de estándares Bookmark. Este método pasó a ser el más utilizado en varios países. Ventajas como su fácil implementación y participación de jueces o expertos, pueden ser la causa del uso masivo que se ha dado a esta metodología.¹¹

Siguiendo este método nosotros realizamos un taller en la ciudad de Quito con nueve representantes académicos de las universidades que ofertan la carrera valorada y un representante de la autoridad sanitaria ecuatoriana.

Al momento del registro de los participantes al taller, se procedió a la firma de un acuerdo de confidencialidad y se procedió a la bienvenida y capacitación sobre la metodología Bookmark.

Como primer paso, dentro de la metodología, los participantes en plenaria discuten sobre el conocimiento y habilidades que debería tener el "estudiante frontera" o "sustentante promedio"

Como segundo paso se entrega a los participantes, para análisis individual, el folleto de ítems ordenados (FIO) el cual contiene una muestra de 56 preguntas ordenadas de menor a mayor dificultad de acuerdo al modelo de Rasch y en función de los resultados obtenidos del examen aplicado. Conjuntamente con los FIO, se entrega a los participantes un documento

que contiene todos los ítems del examen ordenados en base a la dificultad de acuerdo a la Teoría Clásica de los Test (TCT) y a la Teoría de Respuesta al Ítem (TRI).

A continuación se realiza una primera ronda de trabajo denominado “*primer momento*”, en la que en cada una de las mesas, los jueces o especialistas analizan el FIO y discuten sobre qué conocimientos y habilidades debe tener el “*estudiante frontera*” para responder un determinado ítem; así como también la razón por la cual cada ítem es más difícil de contestar que el anterior. Al final del primer momento, cada participante de forma individual ubica la marca entre las dos preguntas según lo indicado anteriormente. La marca se ubica en la pregunta anterior a la considerada como aquella que un sustentante mínimamente cualificado tendría la probabilidad mayor o igual a $\frac{1}{2}$ (un medio) de contestar correctamente.

En la segunda ronda de trabajo, “*segundo momento*” cada mesa inicia con la intervención de cada participante fundamentando las razones por las cuales señaló la primera marca; adicional a este elemento, se presenta al grupo los estadísticos descriptivos de todos los participantes. Con estos dos insumos, cada participante reconsidera o no, la marca establecida en el primer momento. A continuación se sistematiza la información recogida de las marcas de todos los participantes y se presentan los estadísticos descriptivos resultantes del segundo momento y adicional a este elemento, se presenta la habilidad y el porcentaje de habilidad que se requiere para aprobar el examen con el promedio de las marcas del segundo momento y la implicación de esta marca en el porcentaje de estudiantes que aprobarían el examen. Con esta información adicional, en el “*tercer momento*” la discusión se abre a todo el panel fomentando la búsqueda de consensos, luego de ello, cada participante selecciona de manera individual su marca definitiva.

Item Mapping

En la metodología *Item Mapping* desarrollada por Ning Wang en el 2003, la participación de expertos en el área evaluada es esencial; ellos deben identificar un grupo de preguntas o ítems (ordenados por dificultad) donde se espera que un candidato mínimamente competente (MCC), por sus siglas en inglés, conteste una pregunta con una probabilidad de éxito mayor o igual a $\frac{1}{2}$.¹²

Previo a la aplicación del taller se procede a calcular la

dificultad de las preguntas de acuerdo a la teoría de respuesta al ítem, (TRI). Esta dificultad es trasladada a una escala que usualmente tiene un rango entre 70 y 130, según lo señala Wang (2003). Se procede a construir un histograma con rangos de 2 unidades (Figura 1)

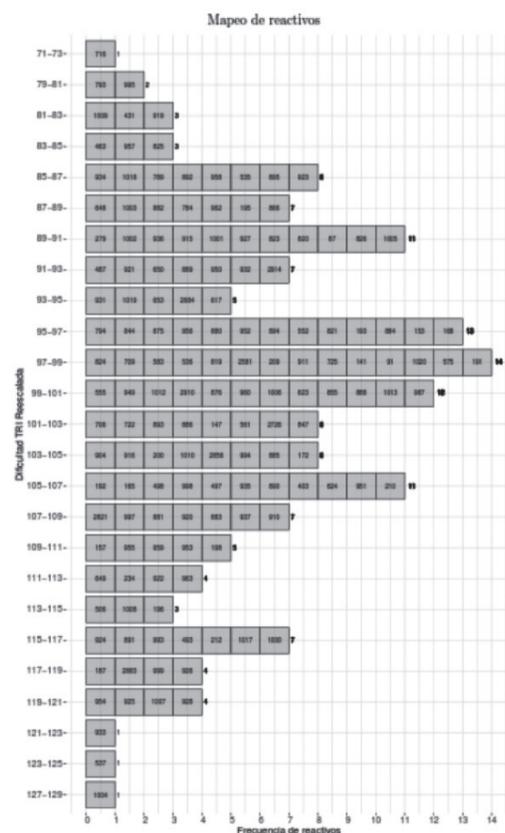


Figura 1. Mapeo de preguntas, método ITEM MAPPING

Este histograma sirve para mostrar una pregunta representante de cada grupo, en orden de dificultad de los grupos, hasta que la mayoría de los jueces o especialistas encuentren un grupo en el cual consideren que un candidato mínimamente competente, MCC, tenga una probabilidad menor a $\frac{1}{2}$ (un medio) de contestar la pregunta correctamente.

En la metodología *Item Mapping*, los talleres se desarrollan de acuerdo al siguiente esquema:¹²

1. Discusión de las características del candidato mínimamente competente.

Para caracterizar MCC, los jueces o especialistas realizan un debate acerca de las competencias y niveles de formación que debería tener, en función de su experiencia como docentes en los últimos niveles de formación.

El objetivo de este primer momento es el de

homologar los criterios que definen a un candidato mínimamente competente, este paso es fundamental para el desarrollo de la metodología.

El candidato mínimamente competente se debe establecer tomando en cuenta las siguientes características:

- El perfil del médico ecuatoriano.
- Las mallas curriculares.
- Morbi-mortalidad.
- Modelo de Atención Integral de Salud (MAIS)

De esta manera se idealizó al candidato mínimamente competente el cual es un sustentante promedio y que servirá de base para relacionar la habilidad con el nivel de dificultad de las preguntas.

2. Se muestran preguntas, representantes de cada grupo, en orden ascendente de dificultad.

Se construye de forma digital, carpetas que contienen las preguntas que se han agrupado de acuerdo con los rangos obtenidos mediante el método probabilístico denominado Modelo de Rasch. Se muestran las preguntas representativas de cada grupo y los especialistas realizan el análisis correspondiente y determinan si un MCC puede o no responder de manera satisfactoria la pregunta presentada. Es decir, se procede a determinar si un MCC podría responder correctamente la pregunta con una probabilidad mayor o igual a 0,5.

3. Se repite el paso anterior hasta obtener que la mayoría de los expertos determinen que el MCC contestaría correctamente con una probabilidad menor o igual a 0,5.

4. Los jueces determinan si el corte cumple con su comprensión conceptual de la competencia mínima y si es necesario iterar el proceso de calificación para ajustar el punto de corte.

En este punto, los especialistas, analizan la pregunta representativa y las demás preguntas de este grupo a fin de determinar si recae en este la dificultad para contestar adecuadamente por parte de un MCC. Al final, se presenta el número de examinados que pasarían con ese punto de corte.

Metodologías Relativas

Una de las metodologías para la determinación de estándar en un examen y probablemente la más antigua, es la que compara directamente entre los resultados de los sustentantes, del tal forma que la aprobación del examen viene dada por una tasa de aprobación establecida previamente.¹⁰

En este sentido, para el análisis se emplea la distribución del porcentaje de respuestas correctas y se determina un porcentaje de aprobación de aproximadamente el 70% (Figura 2). Nótese que esta metodología relativiza los resultados considerando el nivel de aprendizaje de las personas que han rendido el examen.

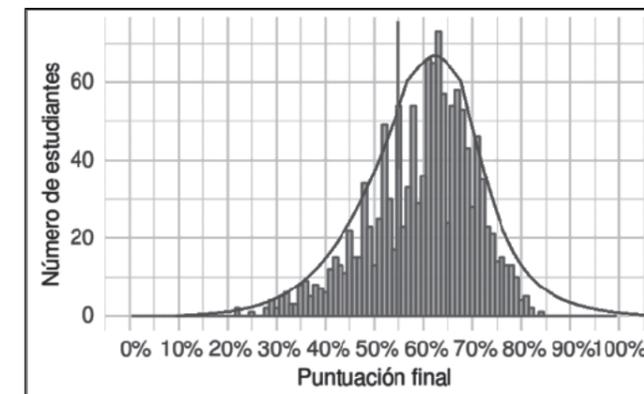


Figura 2. Ejemplo de la distribución del porcentaje de habilidad de un examen. Fuente: Examen de habilitación para el ejercicio profesional, 2017. Elaboración: Ceaces, 2017.

Una vez que se determina el punto de corte (PC), las calificaciones son trasladadas a la escala Ceaces aplicando una función matemática establecida.

$$C = \frac{40}{100 - PC} (PH - PC) + 60$$

Donde:

C: Calificación final

PH: corresponde al porcentaje de habilidad, es decir el número de aciertos sobre el número de preguntas.

PC: es el punto de corte determinado con la metodología normativa. De esta manera, un sustentante obtiene la categoría de satisfactorio con al menos 60/100.

RESULTADOS

Bookmark Standard Setting

En base a lo descrito en la metodología en el primer momento y bajo el esquema propuesto se establecieron los siguientes resultados:

Tabla 1. Estadísticos descriptivos de las marcas del primer momento de dos talleres que desarrollaron la metodología *Bookmark standard setting*

Media	Mediana	Máximo	Mínimo	Rango
41	41	43	39	4

Una vez establecida la segunda marca, se sistematizó la información, que se muestra a continuación:

Tabla 2. Estadísticos descriptivos de las marcas del segundo momento de dos talleres que desarrollaron la metodología *Bookmark standard setting*

Media	Mediana	Máximo	Mínimo	Rango
41	41	43	39	4

Después del análisis estadístico descriptivo de las marcas para determinación del punto de corte se determinan los siguientes resultados que se muestran a continuación:

Tabla 3. Información presentada en el segundo momento de dos talleres que desarrollaron la metodología *Bookmark standard setting*

Segundo Momento	
Número de sustentantes	1317
Punto de corte	41
Porcentaje aprobados	68,2%
Número de aprobados	898
Número de no aprobados	419

En el tercer momento los jueces o especialistas colocaron nuevas marcas, obteniéndose los resultados que se muestra en la tabla 4:

Tabla 4. Resultados finales, obtenidos en el tercer momento de dos talleres que desarrollaron la metodología *Bookmark standard setting*

Media	Mediana	Máximo	Mínimo	Rango	Puntuación estimada	Porcentaje de aprobación
41	41	43	39	4	72,72	68,19

Item Mapping

Con base en este procedimiento, los jueces o especialistas consideraron que en el grupo 13 que corresponde al rango 101-103, se encuentra la pregunta representativa puesto que presentaba la complejidad necesaria para ser contestada adecuadamente por un MCC, por lo tanto, se

procedió a mostrar el resto de las preguntas de este grupo, al final 3 de 4 preguntas fueron consideradas difíciles de contestar para un MCC.

Con estas consideraciones se procedió a realizar los cálculos correspondientes y se presentó el siguiente escenario:

Tabla 5. Resultados obtenidos en taller en el que se desarrolló la metodología *Item Mapping*

DIFICULTAD TRI PREESCALADA	GRUPO	No. DE PREGUNTAS	% DE APROBACIÓN	PUNTUACIÓN ESTIMADA	TOTAL DE SUSTENTANTES
101-103	13	8	4,98 %	106	1317

Metodología Relativa

Los resultados de esta metodología se expresan en números absolutos de acuerdo con la fórmula utilizada y la determinación a priori del porcentaje de aprobados:

Tabla 6. Resultados obtenidos con metodología relativa

Parámetro	Metodología Relativa
Habilidad	79
Porcentaje de Habilidad	54,9

RESUMEN DE RESULTADOS

De las metodologías empleadas se muestra la siguiente tabla donde se puede observar los estándares o puntos de corte determinados y los resultados de aprobación.

Tabla 7. Resultados obtenidos en los tres métodos realizados para obtener puntos de corte que determinen la puntuación estimada y el porcentaje de aprobación de los sustentantes.

Parámetro	Bookmark 1	Item mapping	Relativa
Puntuación estimada	72,72	106	79
Porcentaje de aprobación	68,19	4,98	54,9

DISCUSION

La determinación de estándares y los diferentes métodos utilizados para el análisis posterior a la aplicación de exámenes de alto impacto, garantiza que las decisiones y clasificaciones resultantes del

proceso de evaluación sean más justas y acertadas. Para ello se necesita que los procedimientos utilizados sean válidos y confiables, basados en reglas objetivas que disminuyan en lo posible los elementos de subjetividad que puedan presentarse.

El instrumento de valoración (examen) es un elemento fundamental para la evaluación y se debe garantizar que sea confiable, como base para un posterior análisis de los resultados obtenidos por los sustentantes. La confiabilidad se obtiene en base a la metodología de realización y validación de preguntas entre otros procesos, se describe que una metodología es confiable cuando al repetirse su aplicación se obtienen medidas similares o valores cercanos y se dice que es válido cuando mide específicamente aquello que se propone medir y no otra cosa, esta confiabilidad tiene que ver con cómo mide un instrumento un determinado constructo (qué tan bien lo hace) y la validez se refiere a lo que mide y generalmente orienta sobre qué se puede inferir a partir de los resultados obtenidos.¹³

Asegurar la calidad de las evaluaciones es una de las principales razones que ha dado lugar a la pluralidad de métodos para el establecimiento de estándares y de puntajes de corte.¹⁴

Los métodos basados en los juicios de especialistas, parten del hecho de que no basta saber qué tan bien se desempeñó un sujeto respecto de otro, sino que es indispensable determinar qué tan bien domina un examinado el conocimiento o la habilidad propuesta, esta característica es la que hace referencia a la calidad y al componente subjetivo de la evaluación. A partir de este concepto hemos realizado metodologías absolutas (con la participación de jueces o especialistas de las universidades involucradas en la evaluación y miembros de la autoridad sanitaria ecuatoriana), y metodologías relativas, los resultados demuestran que los métodos basados en el juicio de experto tienen muchas fortalezas, como podemos ver en nuestra experiencia al confirmar la diferencia marcada entre la puntuación estimada que genera el porcentaje de aprobación entre los métodos BOOKMARK STANDARD SETTING (68,19) y el método relativo utilizado (54,9). Se debe tomar en cuenta que estas metodologías tienen algunas limitaciones en relación a la capacidad objetiva para determinar los niveles de desempeño. Por lo tanto, dado que la calidad de la interpretación depende del juicio de jueces o especialistas, la base de estos métodos es garantizar

que los procedimientos sean lo más rigurosos y exhaustivos posibles, el Ceaaces se encuentra afinando estos procesos en cada aplicación por medio de talleres a los especialistas involucrados, para robustecer el proceso, único en el país reduciendo la posibilidad de error y arbitrariedad en los resultados obtenidos.^{15,18}

Tomando en cuenta que los métodos más conocidos y utilizados en el mundo por sus buenos resultados son el método de Angoff (1971)⁵ y el método Bookmark (Lewis et al., 1996; Lewis et al., 1999),⁷ y que métodos como el de Ebel (1972),¹⁶ las modificaciones al de Angoff de Hambleton y Plake (1995)¹⁷ y el de Cooper-Loomis y Bourque (2001),¹⁸ son menos conocidos pero también son empleados, y que de estos métodos el más utilizado en la mayoría de los estados norteamericanos y en diferentes países de Centro América y Latinoamérica como Perú, Chile, Guatemala, Brasil y México en diferentes disciplinas y grados académicos es el método Bookmark standard setting, proponemos considerarlo como un método que puede ser útil en la determinación del punto de corte y estándar en los exámenes aplicados en nuestro medio, considerando además que en el ejercicio presentado en este documento es uno de los que reporta mejores resultados en relación con el índice de aprobación y que esta característica metodológica es similar al resultado satisfactorio encontrado con su implementación en otros estudios.¹⁹

Una de las fortalezas del método Bookmark, es que tiene como principal característica que ha sido diseñado utilizando la Teoría de Respuesta al Item (TRI), lo que otorgan algunos beneficios a los especialistas que analizan los resultados de los exámenes a partir de su estructura, el más importante beneficio es la propiedad de invariancia de los parámetros de los procedimientos de la TRI, una vez establecidos los puntos de corte adecuados, éstos no variarán aunque se modifique o añadan nuevos reactivos al banco de datos, siempre que se utilicen los diseños de equiparación métrica necesarios, además, la TRI permite mapear los ítems según su nivel de dificultad. Es decir, permite calibrarlos para presentarlos ordenados en dificultad a los jueces o especialistas que participarán en el establecimiento de puntos de corte, otro beneficio es el cuadernillo, el que permite a los especialistas tener una visión global del grado de complejidad de la prueba y ayuda a que focalicen su atención en aquellos ítems particularmente sensibles para los estudiantes

límites, facilitando la determinación del punto de corte basados también en niveles de competencia.²⁰

En conclusión, el establecimiento de estándares tiene por finalidad generar procesos de evaluación confiables y válidos que permitan conocer los resultados de aprendizaje y las habilidades que deben dominar los sustentantes en los diferentes niveles de desempeño, esto busca favorecer el diseño y la implementación de políticas tanto educativas como en el área de salud en el caso ecuatoriano, a fin de que determinados aprendizajes comunes a cada área disciplinar y a cada grado estén garantizados, El uso de estas metodologías en las aplicaciones periódicas desarrolladas por el Ceaces promueve una mejora de la calidad educativa de la población evaluada y la coordinación multidisciplinaria entre los diferentes órganos de control.

BIBLIOGRAFIA

1. Melchor Sánchez-Mendiola, Laura Delgado-Maldonado. Exámenes de alto impacto: Implicaciones educativas. *Investigación en Educación Médica*. 2017; 6(21): 52-62.
2. Wang, N. Use of the Rasch IRT Model in Standard Setting: An item-mapping method. *Journal of Educational Measurement*. 2003. 40(3) 231-253
3. Mendoza Ramos, Arturo. La validez en los exámenes de alto impacto: Un enfoque desde la lógica argumentativa. *Perfiles educativos*, 2015; 37(149):169-186
4. Dwyer, T., Wright, S., Kulasegaram, K. M., Theodoropoulos, J., Chahal, J., Wasserstein, D., ... & Ogilvie-Harris, D. How to set the bar in competency-based medical education: standard setting after an Objective Structured Clinical Examination (OSCE). *BMC medical education*. 2016; 16(1), 1.
5. Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement*. 1971; 508-600. Washington, DC: American Council on Education
6. Lewis, D. M., & Green, D. R. The validity of performance level descriptors. Paper presented at the Council of Chief State School Officers National Conference on Largescale Assessment, Phoenix, AZ. 1997.
7. Lewis, D. M., Mitzel, H. C., Green, D. R. & Patz, R. J. (1999). The bookmark standard setting

procedure. Monterey, CA: McGraw-Hill.

8. Herrera Ortiz, M., Benavides Posadas, D., & Monroy Cazorla, L. (2009). Establecimiento de estándares en un examen criterial. Cuaderno Técnico 3. Méjico, D.F.: Centro Nacional de Evaluación para la Educación Superior.
9. Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, M. (2012). The Bookmark Standard Setting Procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods and innovations* (pp. 225-253). New York, NY: Routledge.
10. Cizek, G. J., Bunch M.B. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Sage publications. Thousand Oaks, London, New Delhi. 2007
11. MacCann R.G, and Stanley G. The Use of Rasch Modeling to Improve Standard Setting. *Practical Assessment, Research and Evaluation*. 2006. 11(2):1-17.
12. Wang N, Wiser R.F, Newman L.S. Use of the Rasch IRT Model in Standard Setting: An item-mapping method. *Journal of Educational Measurement*. 2003;231-253.
13. Kingston, N.M., Kahl, S.R., Sweeney, K.P., & Bay, L. Setting performance standards using the Body of Work method. In G.J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods and Perspectives*. 2001; 219-247. Mahwah, NJ: Erlbaum.
14. Jornet Meliá, J. M. & González Such, J. Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*. 2009;16, 103-123.
15. Hambleton, R. H. Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. 2001; 89-116. USA: Taylor & Francis.
16. Ebel, R. L. Content standard test scores. *Educational and Psychological Measurement*. 1972; 22; 15-25.
17. Hambleton, R. K., & Plake, B. S. Using an extended Angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education*. 1995; 8; 41-56.
18. Cooper-Loomis, S., & Bourque, M. L. From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. 2001;175-217.

Mahwah, NJ: Erlbaum.

19. Nellhaus, J. M. States with NAEPLike Performance Standards. In M. L. Bourque & S. Byrd (Eds.) *Student performance standards on the National Educational Assessment of Educational*

Progress: Affirmation and improvements. 2000; 99-130. Washington, DC.

20. García, P. E., Abad, F. J., Olea, J., & Aguado, D. A new IRT-based standard setting method: Application to eCat-Listening. *Psicothema*. 2013;25(2);238-244.