

**Evaluación de los estudiantes:  
las prácticas institucionales y docentes  
vistas desde el marco de estándares internacionales**

*Assessment of students achievement: Institutional and teachers practices seen within the framework of international standards for educational evaluation*

**Katherina Edith Gallardo Córdova**

**Arcelia Ramos Monobe**

**Dulce Camacho Gutiérrez**

**María Eugenia Gil Rendón**

Tecnológico de Monterrey, México.

**RESUMEN**

El propósito de esta investigación fue analizar las prácticas de evaluación del aprendizaje desde el marco de los estándares internacionales de calidad: factibilidad, ética, precisión y utilidad, enunciados por el Comité Conjunto sobre Estándares de la Evaluación Educativa (Joint Committee on Standards for Educational Evaluation, JCSEE), con la finalidad de encontrar las fortalezas y áreas de oportunidad en las prácticas del profesorado y las instituciones educativas en diferentes contextos. Se optó por un estudio exploratorio debido a la escasa indagación sobre el tema en el contexto latinoamericano. Participaron 347 profesores y directivos de siete países. Se aplicó un cuestionario con escala tipo Likert, diseñado con base en los estándares internacionales de evaluación: ética, factibilidad, utilidad y precisión. Los resultados dejan ver que, desde la perspectiva de profesores y directivos, las prácticas de evaluación en general se orientan al cumplimiento de los estándares internacionales. Sin embargo, aspectos como la comunicación de resultados a los interesados y públicos diversos, la participación de estudiantes y padres en la toma de decisiones y ciertos estándares sobre ética en la evaluación no están del todo afianzados en la cultura de evaluación de los participantes. La visión integral sobre las prácticas indagadas, enmarcada en lo estándares, permite formular preguntas y retos de investigación a futuro que se vislumbran como componentes cruciales en el fortalecimiento de la cultura de evaluación en escenarios de educación presencial y a distancia.

**Palabras clave:**

Estándares, normas, profesores, estudiantes, evaluación del aprendizaje.

**ABSTRACT**

*The purpose of this research was to analyze the practices of learning evaluation in the frame of international quality standards: feasibility, ethics, accuracy and usefulness enunciated by the Joint Committee on Standards for Educational Evaluation (JCSEE) in order to find the strengths and weaknesses in the practices of educational practices in different contexts. The standards JCSEE postulate are: Ethics, feasibility, utility and accuracy. An exploratory design was chosen due to insufficient research that exists on this topic in the Latin American context. A total of 347 teachers and principals participated in this study across seven countries. A Likert scale questionnaire was designed based on international standards of evaluation. The results reveal that from the perspective of teachers and principals, assessment practices are generally oriented towards meeting international standards. However, it is inferred that aspects such as communication of results to stakeholders as well as involving students and parents in decision-making are not fully entrenched in the assessment culture of participants. This problem also aroused in certain standards referred to ethics. The holistic view this research offers allows researchers to raise questions for educational investigation upon critical components that could strength evaluation practices in a near future for both face to face and distance education.*

**Keywords:**

Standards, regulations, teachers, students, learning evaluation.

## INTRODUCCIÓN

Las prácticas de evaluación del aprendizaje, frente a las exigencias de la sociedad y la economía del siglo XXI, han sido objeto de estudio y reflexión en las últimas dos décadas a nivel global. El tema ha llevado a organismos internacionales a emitir recomendaciones a nivel político, conceptual y técnico con afán de mejorar los procesos de evaluación y alcanzar así niveles óptimos de calidad. Algunas de las recomendaciones más importantes en materia de evaluación del aprendizaje se enfocan al desarrollo y adopción de nuevas métricas que faciliten el manejo e interpretación de resultados que permitan integrar la información; el apoyo permanente que debe otorgarse a las comunidades de práctica en materia de cambio y mejora de los sistemas de evaluación; el fomento al intercambio de conocimientos, lecciones aprendidas y ejemplos de prácticas eficaces; y el involucramiento de la sociedad civil para ayudar en la medición y así mejorar el aprendizaje (BID, 2011; OCDE, 2014; UNESCO, 2013 y 2014).

En este mismo orden de esfuerzos, el JCSEE ha realizado durante varias décadas procesos de postulación y aprobación de propuestas para el establecimiento de estándares que conlleven a la generación de buenas prácticas educativas en evaluación. Si bien es cierto que los estándares propuestos por el JCSEE están a disposición pública, es muy probable que su difusión en América Latina sea escasa o casi nula. No obstante, por la importancia que el tema está cobrando en los últimos años, juzgamos valioso trabajar en un estudio que coadyuvara en su comprensión y difusión en el campo educativo, como algunas investigaciones en el tema lo han venido realizando (Escudero, 2013; Kimball & Milanowski, 2009).

La intención de este estudio es contribuir con una primera aproximación que permita iniciar conversaciones sobre cómo docentes de diferentes instituciones y niveles educativos juzgan sus prácticas actuales en evaluación del aprendizaje en relación con lo que los estándares internacionales estipulan.

## MARCO TEÓRICO

Las prácticas de evaluación del aprendizaje son tan diversas como amplios son los horizontes que planes de estudio y programas visualizan para llegar a la adquisición de conocimientos y desarrollo de habilidades. En este sentido, actualmente la diversidad en cuanto a la conceptualización e interpretación de cómo realizar evaluación del aprendizaje de cada docente genera una fuente inagotable de justificaciones que respaldan su práctica (Álvarez, 2001; Martínez, 2013). Esto lleva a pensar que la indagación científica del fenómeno evaluativo se vuelve cada vez más compleja, pero a la vez necesaria cuando la intención que se persigue es comprender las prácticas del profesor en el aula y cómo orientarlo para que su labor en evaluación, en efecto, coadyuve al desarrollo de las competencias de sus estudiantes.

Esto concuerda con lo que Escudero (2013) sostiene sobre tomar ciertas consideraciones al realizar investigación evaluativa. Ésta se inserta en un ambiente y, por tanto, el contexto también influye en la forma en que se evalúan los programas, los sistemas y las instituciones. La investigación sobre evaluación se efectúa en el marco de ciertas intenciones subyacentes sobre su papel en el proceso enseñanza-aprendizaje y es impensable que se lleve a cabo en un contexto neutro.

En la actualidad, el cambio más trascendente se relaciona con la intención de pasar de ser una evaluación *del* aprendizaje a una evaluación *para* el aprendizaje. Esto, más que ser un juego de palabras, coloca al campo de la investigación en evaluación en otra dimensión: de considerar la evaluación como un proceso para la emisión de juicios del aprendizaje *a posteriori* a transformarse en un proceso entendido como activo y significativo, en tanto que acerca *a priori* a los estudiantes a comprender su proceso de aprendizaje y hacer lo necesario para mejorar su aproximación a las metas educativas (Moreno, 2012). Este giro implica el sostener las prácticas basadas en marcos de referencia que provoquen el cambio de prácticas de evaluación en el aula.

Cada vez con más fuerza se ha venido trabajando en estos marcos de referencia a nivel internacional, y consisten en la promulgación y aceptación de ciertos criterios que expresan lo que se debe hacer para sostener mejores procesos de evaluación desde su planeación hasta su valoración. A éstos se les denomina estándares y son un principio de calidad, acordado por un comité de expertos. Las declaraciones de estándares para la evaluación se han publicado desde mediados del siglo XX y se han revisado de manera constante con el fin de enriquecerlos (Yarbrough, Shulha, Hopson & Caruthers, 2011). Así, los estándares son orientaciones, guías y lineamientos que armonizan los sistemas de evaluación del alumnado

y las prácticas evaluativas de los profesores y las instituciones (Cabra, 2008; Ravitch, 1996). Ravitch (1996) afirma que los estándares buscan indicar el tipo de logro al que corresponde determinado esfuerzo, además de promover la cooperación entre profesores al tiempo que estimulan el mejoramiento de la instrucción y se enmarcan las características que la evaluación debe respetar.

Uno de los grupos académicos en el tema de estándares con una constante actividad de revisión y consultoría a nivel internacional es el JCSEE. Este organismo fue creado para cubrir una necesidad en cuanto a la indagación sobre el uso y la aplicación de estándares en contraposición al empleo exclusivo de pruebas (test) para medir el aprendizaje. En 1974, representantes de la Asociación Americana de Psicología, de la Asociación Americana de Investigación Educativa y el Consejo Nacional de Medición Educativa decidieron reunirse para trabajar el tema. Se determinó que la evaluación no solamente se hacía con los estudiantes y se podía ir hacia la inclusión de otros elementos inmersos en el proceso educativo y que, por tanto, debían generar estándares propios para guiar la evaluación más allá de las pruebas.

Así, la primera versión de estándares para la evaluación de programas educativos, proyectos y materiales fue publicada en 1981. Después de dicha publicación, este grupo reflexionó que había que mantenerlos, revisarlos y mejorarlos a lo largo del tiempo. Desde entonces, el JCSEE revisa y publica sus obras cada cinco años. Asimismo, capacita a los responsables políticos, evaluadores y educadores en el uso de las normas y funge como centro de intercambio y discusión de literatura sobre estándares de evaluación. Los estándares planteados por la JCSEE son utilidad, factibilidad, ética y precisión (Sanders, 1999).

De acuerdo con Yarbrough, Shulha, Hopson et al. (2011), los estándares de utilidad tienen como objetivo que los usuarios encuentren valor, tanto en los procesos como en los productos referentes a la evaluación, para satisfacer sus necesidades. En cuanto a factibilidad, este tipo de estándares buscan marcar la medida en que la evaluación debe ser eficiente y eficaz o en una pregunta: ¿es posible que la evaluación cumpla con aquello para lo que fue ideada? Cuando se cita la ética en los estándares, se hace alusión a lo correcto, legal, justo, equitativo, y apropiado. Por último, los estándares de precisión buscan brindar la veracidad y la confiabilidad en los procesos de evaluación, sus afirmaciones y conclusiones, sobre todo en la emisión de interpretaciones y juicios de valor.

Uno de los primeros estudios en el tema de estándares de evaluación fue el realizado por Shapiro (1973), el cual giró en torno a valorar los procesos de evaluación y sirvió para comprender qué tanto el intelecto y la emoción intervienen en el proceso de aprendizaje. A partir de éstos se determinaron los alcances que pueden llegar a tener las pruebas (o *tests*) como instrumentos de evaluación. El estudio contó con seis grupos de primer año de primaria. El proceso de indagación se enriqueció con el uso de guías de observación, entrevistas, información sobre los factores familiares y registro de calificaciones. Los hallazgos marcaron una disparidad considerable entre las respuestas emitidas dentro del salón de clases y las respuestas en pruebas individuales. Esto llevó a reconsiderar la racionalidad de las pruebas utilizadas. Se concluyó que las diferencias encontradas entre niños con antecedentes y contextos similares podrían deberse a fallas en cuanto a la adecuación, utilidad, representatividad y aplicación de las pruebas. Esto, sin duda, dio pie a una serie de cuestionamientos sobre los estándares en los cuales se apoyaba el proceso de evaluación.

Por su parte, Cabra (2008) exploró las percepciones de los estudiantes y sus profesores en cuanto a la calidad de las evaluaciones. El estudio tuvo lugar en dos centros de educación privada en España. Contó con respuestas de 311 estudiantes de secundaria obligatoria y bachillerato, y utilizó un cuestionario de 32 ítems sobre los cuatro principios de los estándares: 8 ítems indagaron sobre el principio de ética, 9, sobre utilidad, 4, sobre viabilidad y 11, sobre precisión. Los resultados integraron un conjunto de ocho factores ordenados del más al menos relevante:

- Usos, que tienen que ver con saber qué tanto han aprendido y cuáles son sus áreas de oportunidad.
- Revisión del proceso, relacionada con las condiciones y actividades que favorecen la mejora del aprendizaje del estudiante.
- Información como retroalimentación para que el alumno conozca cuáles fueron sus resultados, con base en qué se emitieron los juicios, así como su oportunidad y transparencia.
- Orientaciones que brinda la evaluación para que los estudiantes y sus familias puedan identificar métodos de evaluación, procedimientos y criterios.
- Legitimidad, reflejada en el respeto a sus derechos y el cuidado de su integridad durante el proceso de evaluar.

- Claridad de la información, en relación con los objetivos de aprendizaje y actividades que conforman el proceso.
- Garantía, es decir, el grado de identificación y comprensión de la manera en que los aprendizajes han sido logrados, el ajuste de las prácticas de enseñanza a las situaciones observadas y el apoyo de las familias para que, tanto ellas como la escuela, compartan la misma visión sobre metas y fines de la evaluación.
- Objetividad, de lo aprendido y de las áreas de oportunidad.

La importancia de la comprensión de los procesos de evaluación en las escuelas para su mejor orientación es explicada por Kimball y Milanowski (2009), quienes trabajaron con directores educativos. A ellos se les solicitó dar a conocer su percepción sobre cómo sus profesores seleccionan instrumentos de evaluación y emiten resultados para determinar el desempeño de sus estudiantes. Este estudio se realizó en ocho escuelas. En cuatro de ellas durante dos años existió consistencia en los resultados de evaluación y en las otra cuatro, en el mismo periodo, se presentó inconsistencia en el mismo rubro.

Los hallazgos, independientemente de la clasificación inicial de las escuelas, arrojan variaciones sustanciales en cuanto a cómo los directores y sus profesores toman decisiones para conducir los procesos de evaluación y las formas en que éstos son juzgados como apropiados y válidos según necesidades educativas y contexto escolar. Esta inconsistencia, de acuerdo con los autores, podría explicar las razones por las cuales los índices de validez y confiabilidad en cuanto a los resultados de las evaluaciones en general son bajos. Las recomendaciones del estudio apuntan a preparar mejor a directivos y profesores en las decisiones sobre elección de mecanismos de evaluación considerando la validez y confiabilidad de los instrumentos, así como la incidencia que deben tener los líderes educativos en este tema.

Otro estudio, en este caso de tipo descriptivo, lo realizó Peterson, Kelly y Caskey (2006) en el campo del estándar ético desde la perspectiva de la evaluación entre docentes. La revisión bibliográfica de los organismos involucrados y reguladores de esta práctica evaluativa y el levantamiento de información entre docentes permitió hallar un total de 16 actividades relacionadas con la ética en la evaluación de colegas. Entre dichas actividades se juzga como de mayor relevancia:

- Conocimiento del rol integral del docente, el manejo confidencial de información (recopilar, utilizar, divulgar, custodiar).
- Apertura hacia el ser humano y la resolución de conflictos.
- Participación en su propia evaluación.
- Preparación profesional requerida para enfrentar el trato con personas cuando se emiten juicios.
- Dar crédito a las ideas de otros.

Se concluyó que aunque se identificaron los aspectos esenciales en el ámbito de la ética en torno a la evaluación de pares y autoevaluación, emergió la necesidad de guiarse y conocer más sobre los estándares para realizar los procesos de emisión de una autoevaluación y evaluación entre pares de la mejor manera posible.

## **MÉTODO**

### **Diseño**

Realizamos un estudio exploratorio en instituciones educativas de varios países con el objetivo de contar con una primera aproximación de la percepción de docentes en ejercicio sobre las propias prácticas de evaluación del aprendizaje, así como algunas cuya responsabilidad recae en la institución donde labora. Estas prácticas se analizaron a la luz de estándares internacionales declarados por el JCSEE. Consideramos apropiado llevar a cabo un estudio exploratorio dada la escasa producción de artículos científicos en los que actualmente se aborda el tema, en particular en contextos latinoamericanos (Hernández, Fernández-Collado y Baptista, 2010; Valenzuela y Flores, 2012; Pérez, Galán y Quintanal, 2012).

A través de la página de Redalyc, base de datos de publicaciones en América Latina y el Caribe, aparecen dos artículos relacionados con estándares de evaluación publicados en los últimos cinco años (Cabra, 2008 y Escudero, 2013), mientras que en el IRESIE, base de datos de la Universidad Nacional Autónoma de México (UNAM), no aparecen artículos sobre el tema de estándares de evaluación, aunque sí hay 64 artículos relacionados con estándares curriculares y propios de diversas disciplinas entre ciencias exactas y naturales. Lo anterior permite inferir que es un tema de reciente interés.

## Marco contextual

La conducción de esta investigación se dio dentro de un curso de posgrado en línea cuyo objetivo se centra en el desarrollo de competencias para desempeñarse mejor en materia de evaluación del aprendizaje. Este curso forma parte del programa en modalidad virtual de la maestría en Educación de una universidad del noreste de México. Los alumnos que se inscribieron en esta materia son de diversas nacionalidades.

Una de las actividades de inicio del curso buscó acercar aspectos teóricos a la realidad contextual, lo cual implicó que cada estudiante realizara un proceso de diagnóstico sobre las prácticas de evaluación de profesores y directores para comprender el fenómeno desde la perspectiva de estándares internacionales en las diferentes instituciones educativas en las cuales los alumnos laboraban o participaban en alguna forma.

## Participantes

Los participantes fueron profesores y directivos de mandos medios y altos que al momento de la encuesta estuvieran trabajando en una institución educativa. Dado que el motivo de este estudio son las prácticas de evaluación en el marco de los estándares internacionales, incluimos a profesores de instituciones públicas y privadas de diferentes niveles educativos (desde educación básica hasta profesional). El total de profesores y directivos participantes fue de 347, quienes ejercen en diferentes países: 134 en México, 183 en Colombia, 9 en Costa Rica, 12 en Ecuador, 3 en Guatemala, 3 en Perú y 3 en Estados Unidos de Norteamérica. En total, sumaron 120 instituciones participantes.

Después de la aplicación, realizamos la concentración de una sola base de datos para su análisis. Consideramos conveniente incluir profesores de diferentes niveles educativos y tipos de instituciones, dado que el objetivo era tener una primera aproximación a las prácticas de evaluación en el marco de estándares en general. Hay que destacar que todos ellos expresaron aprobación de uso de datos a través de cartas de consentimiento que personalmente firmaron y enviaron.

## Instrumentación

Utilizamos una encuesta con escala tipo Likert para estimar cuán frecuentemente se presentan en el aula las prácticas de evaluación del aprendizaje que mejor se acercan a cumplir con los estándares internacionales. Este instrumento denominado *Cuestionario para emitir un diagnóstico sobre prácticas de evaluación del aprendizaje en escenarios educativos* (ver anexo) fue construido por las investigadoras tomando en cuenta cada uno de los estándares e indicando la actividad que se demanda en cuanto a evaluación del aprendizaje se refiere. Por tal motivo, las categorías son las mismas que establece el JCSEE: ética, factibilidad, utilidad y precisión. Asimismo, se verificó la existencia de otros instrumentos con similar objetivo para fortalecer su construcción (Clipa y Duta, 2010).

El cuestionario se estructuró con base en las siguientes categorías y sus respectivos ítems:

### Ética

1. En mi institución los lineamientos de evaluación del aprendizaje se apegan a la misión, la visión y los principios que rigen el quehacer educativo.
2. En mi institución existe un manual sobre procedimientos de evaluación del aprendizaje, el cual mis colegas profesores y yo seguimos.
3. Si solicito revisar los antecedentes en cuanto a calificaciones y retroalimentación sobre el desempeño de mis estudiantes en periodos anteriores, esta información me es proporcionada sin problema.
4. Las evaluaciones que hago para estimar el aprendizaje se conducen en un marco de respeto y protección de la dignidad de los estudiantes.
5. Las evaluaciones que hago para estimar el aprendizaje se enmarcan en el reglamento académico de la institución.
6. Las evaluaciones que realizo tienen como propósito estimar qué tanto los estudiantes han logrado los objetivos de aprendizaje o las competencias establecidas para la disciplina y el nivel educativo.
7. Las evaluaciones que realizo permiten diferenciar claramente cuáles son las fortalezas y debilidades de cada estudiante.

### Utilidad

8. Al aplicar procesos de evaluación los estudiantes están conscientes de que éstos apoyan la construcción de conocimiento o desarrollo de habilidades.
9. Comunico los resultados de evaluación con el fin de tomar decisiones y acciones que coadyuven al logro de las metas educativas.

10. En mi institución están definidos quiénes son los usuarios de evaluación: los estudiantes, los profesores, el área psicopedagógica, la dirección y el padre o tutor de cada estudiante.
11. En mi institución están definidos los usos que se le deben dar a los resultados de las evaluaciones del aprendizaje.
12. Realizo reportes con base en los resultados de las evaluaciones del aprendizaje en los periodos que indica la institución.
13. En los reportes que realizo para cada estudiante o el grupo de estudiantes, cuido la delimitación de objetivos de evaluación, interpretación de los resultados y recomendaciones.
14. Considero que tengo la preparación suficiente para conducir procesos de evaluación formativa y sumativa.
15. Considero que los profesores de la plana docente cuentan con preparación suficiente para conducir procesos de evaluación formativa y sumativa.
16. Los estudiantes han sido informados sobre los valores que sustentan los procesos de evaluación, entre los que destacan, por ejemplo, la equidad en las aplicaciones y la discreción en cuanto a la comunicación de resultados.
17. Al recibir los resultados de evaluación, los estudiantes también reciben información sobre las estrategias a seguir para mejorar el alcance de los conocimientos y desarrollo de habilidades que se estiman pertinentes.
18. Los profesores elaboran reportes de seguimiento de los estudiantes cuya evaluación denotó que están en riesgo de no alcanzar niveles satisfactorios de desempeño.

### **Factibilidad**

19. Planeo las evaluaciones tomando los objetivos de aprendizaje, lo cual me permite plantear procesos que estimarán qué tan cerca están los estudiantes de alcanzarlas.
20. Considero que los resultados que emito son fidedignos, sin alteraciones de ninguna índole.
21. Tomo en cuenta las opiniones de los estudiantes al momento de decidir sobre los procesos de evaluación.
22. Considero las opiniones de los padres o tutores al momento de tomar decisiones sobre los procesos de evaluación.
23. La institución provee materiales y condiciones para planear los procesos de evaluación de forma efectiva y segura.
24. La institución provee materiales y condiciones para efectuar los procesos de evaluación de forma efectiva y segura.

### **Precisión**

25. Conozco las escalas de evaluación que se utilizan para determinar el nivel de desempeño de los estudiantes.
26. Los objetivos de aprendizaje o competencias declaradas me dicen claramente cuál es el nivel de desempeño que espero de mis estudiantes.
27. Los objetivos de aprendizaje o competencias declaradas me permiten justificar el tipo de actividades de evaluación formativa y sumativa que diseño y aplico.
28. Controló las condiciones que podrían afectar la aplicación de procesos de evaluación: excesivo calor o frío en el aula, hora de la aplicación, tiempo de trabajo en la evaluación, etcétera.
29. Controló las condiciones que podrían afectar la resolución de las actividades por parte de los alumnos como, por ejemplo, el plagio, la sustitución de identidad, la consulta de material no autorizado mediante el proceso, etcétera.
30. Las calificaciones y observaciones derivadas de procesos de evaluación se resguardan para futuras consultas.
31. Realizo procesos de análisis y mejora de los mecanismos de evaluación: actualización de tablas de especificaciones, bancos de preguntas, análisis estadísticos de resultados de exámenes, etc.
32. La institución prevé dentro de los procesos de evaluación una etapa posterior a la aplicación en la cual los colegas y yo revisamos y corregimos aquellos que necesiten mejorar.
33. La institución realiza análisis de resultados de evaluación para estimar alcances y necesidades de seguimiento.
34. En la institución se realizan procesos de valoración de los mecanismos de evaluación del aprendizaje para determinar necesidades de mejora y cambio.
35. Además de los reportes que yo emito, la institución realiza otros reportes para públicos diferentes al estudiante: autoridades educativas, padres de familia, etcétera.

El instrumento se sometió a un proceso de validez de constructo y contenido. En cuanto a la confiabilidad, el instrumento sostiene niveles de 0.82 según el análisis de alfa de Cronbach obtenido. La escala se construyó con cinco valores que conformaron una escala, a lo largo de las 35 afirmaciones que constituyen el instrumento. La dirección de cada una de las afirmaciones fue favorable. Los valores de la escala para cada una de las categorías fueron las siguientes: 1) nunca, 2) casi nunca, 3) a veces, 4) casi siempre y 5) siempre. Sólo en dos de las 35 afirmaciones se hizo una variación en la escala para expresar acuerdo o desacuerdo con la afirmación en vez de frecuencia, dado que partieron de estándares que involucran directamente las capacidades y preparación del profesor y no tanto sus prácticas.

Así, los ítems 14 y 15 de la escala utilizada fueron los siguientes: 1) totalmente en desacuerdo, 2) en desacuerdo, 3) ni de acuerdo ni en desacuerdo, 4) de acuerdo y 5) totalmente de acuerdo. Optamos por incluir un valor en la escala que permitiera expresar desconocimiento, denominada 6) no tengo información, como ayuda para categorizar mejor las respuestas en dado caso que alguna de las prácticas apegadas a los estándares de evaluación no estuviera del todo entendida o no fuera familiar para el docente. Esta información también fue apreciada para comprender mejor el fenómeno desde la delimitación del desconocimiento en la práctica.

#### *Procedimiento*

El protocolo de aplicación fue similar en todos los escenarios. Primero, realizamos un proceso de firma de una carta de consentimiento. Luego, entregamos el instrumento a los participantes. Los aplicadores dieron tiempo suficiente para escuchar las indicaciones, leer el instrumento y contestar. Los aplicadores utilizaron tecnología de documentos compartidos en coordinación con las investigadoras, lo cual permitió integrar una sola base de datos. Una vez completa la base de datos, la sometimos a procesamiento estadístico con apoyo del programa Excel. Todas las respuestas fueron procesadas de acuerdo como deben tratarse las bases de datos que se forman a partir de la aplicación de cuestionarios. Sólo la opción 6 (no tengo información) fue retirada del análisis, aunque se trató como valor de frecuencia.

## **RESULTADOS**

A continuación, presentamos los resultados de la aplicación de los instrumentos a los diferentes grupos de forma general y, posteriormente, clasificamos los resultados por cada tipo de estándar: ética, utilidad, factibilidad y precisión. Al final, mostramos los resultados de la consistencia interna del instrumento.

### **Ética**

La tabla 1 contiene los resultados obtenidos en todos los grupos respecto al estándar denominado ética.

**Tabla 1.** Resultados del cuestionario en la sección correspondiente a la ética.

Ítem	Promedio	Desviación estándar
# 4	4.86	0.44
# 6	4.72	0.52
# 5	4.64	0.71
# 7	4.33	0.71
# 3	4.28	1.07
# 1	4.25	0.86
# 2	3.61	1.44

Es preciso destacar algunos resultados sobre el estándar de ética; por ejemplo, en el ítem # 4, que indaga qué tanto durante la evaluación se cuidan aspectos como respeto y dignidad del estudiante, se alcanzó el mayor promedio y la desviación estándar más reducida (prom. 4.86, desv. est. 0.44). Un segundo resultado importante es el del ítem # 3, asociado a la revisión de los antecedentes de los estudiantes sobre su rendimiento académico; este ítem, aunque puntuó alto en el acuerdo de ejecución, presentó una desviación de un punto, lo cual representa una variabilidad considerable (prom. 4.28, desv. est. 1.07). El ítem # 2, por otro lado, es el que menor puntuación logró y mayor variabilidad; esto se puede interpretar como una debilidad en cuanto a la existencia y uso de manuales que establezcan normas para el proceso evaluativo (prom. 3.61, desv. est. 1.44).

## Utilidad

La tabla 2 contiene los resultados obtenidos en todos los grupos en este estándar.

**Tabla 2.** Resultados del cuestionario en la sección correspondiente a la utilidad.

Ítem	Promedio	Desviación estándar
# 9	4.54	0.74
# 12	4.46	0.97
# 14	4.30	0.80
# 10	4.28	1.01
# 11	4.18	1.06
# 13	4.16	0.97
# 16	3.96	1.13
# 8	3.93	0.98
# 18	3.90	1.10
# 15	3.83	0.97
# 17	3.82	1.10

De los resultados obtenidos en cuanto a utilidad, destaca que el ítem # 9, relacionado con la comunicación de resultados y la toma de decisiones y acciones para favorecer el logro de metas educativas, fue considerado con más alta frecuencia en la práctica, seguido por el # 12, sobre la emisión de reportes de evaluación (prom. 4.54, desv. est. 0.74 y prom. 4.46 y desv. est. 0.97). En lo concerniente a las prácticas que se perciben como menos realizadas en el proceso evaluativo, se encuentran las que tienen que ver con la preparación del personal docente para conducir procesos de evaluación, así como con la entrega de resultados y estrategias para que los estudiantes alcancen a desarrollar habilidades y conocimientos que los impulsan a lograr las metas educativas (prom. 3.83, desv. est. 0.97 y prom. 3.82, desv. est. 1.10).

## Factibilidad

La tabla 3 muestra los resultados obtenidos en todos los grupos en este estándar.

**Tabla 3.** Resultados del cuestionario en la sección correspondiente a la factibilidad.

Ítem	Promedio	Desviación estándar
# 20	4.72	0.63
# 19	4.48	0.76
# 24	3.83	1.09
# 23	3.77	1.11
# 21	3.57	1.12
# 22	3.02	1.30

Los resultados relacionados con la factibilidad destacan con mayor frecuencia los correspondientes a los ítems # 20 y 19, acerca del uso de resultados fidedignos, sin alteraciones de ninguna clase, y la planeación de las evaluaciones con base en los objetivos de aprendizaje. En ambos casos, el valor de la desviación se considera de mediano alcance (prom. 4.72, desv. est. 0.63 y prom. 4.48, desv. est. 0.76). Sin embargo, prácticas como la del ítem # 21 y del 22, respecto a la consideración de la opinión de padres y estudiantes sobre aspectos de la evaluación, se consideran acciones que rara vez se efectúan, las cuales también obtuvieron desviaciones estándar de alto valor.

## Precisión

La tabla 4 presenta los resultados obtenidos en todos los grupos en este estándar.

**Tabla 4.** Resultados del cuestionario en la sección correspondiente a la precisión.

Número de ítem	Promedio	Desviación estándar
# 25	4.74	0.66
# 26	4.32	0.83
# 30	4.30	1.04
# 27	4.29	0.80
# 29	4.20	0.96
# 31	3.81	1.04
# 33	3.77	1.19
# 34	3.75	1.12
# 35	3.58	1.36
# 28	3.42	1.19
# 32	3.38	1.32

En cuanto a la precisión, en la tabla 4 apreciamos que los docentes aseguran conocer las escalas de evaluación con las que llevan a cabo los procesos de valoración (ítem # 25) y afirman que los objetivos de aprendizaje son suficientemente claros para determinar los mecanismos de evaluación que los conducirán a medir los alcances de los estudiantes (ítem # 26). Ambos ítems se declararon con alta frecuencia y desviación estándar de mediano valor (prom. 4.74, desv. est. 0.66 y prom. 4.32 y desv. est. 0.83). No obstante, manifestaron a través del ítem # 28 tener poco control sobre ciertas condiciones ambientales, de tiempo y espacio cuando aplican evaluaciones, así como escasa oportunidad de mejorar estrategias o instrumentos que requieran ajustes a través del ítem # 32 (prom.3.42, desv. est. 1.19 y prom. 3.38, desv. est. 1.32). En ambos casos, la desviación es considerada alta. Otro resultado que debemos destacar corresponde al ítem # 35. Aunque el promedio de puntuación no es el más bajo, la desviación estándar es la más alta obtenida a lo largo del instrumento, lo cual puede indicar que la práctica de emitir reportes de las evaluaciones para públicos diferentes a estudiantes podría presentarse como irregular.

Respecto al valor 6 de la escala Likert, utilizado para que los participantes expresaran su desconocimiento sobre si una práctica se efectuaba o no, los resultados arrojaron que la afirmación con mayor frecuencia de desconocimiento es el ítem # 35, vinculado al estándar de precisión, con un total de 46 incidencias, lo cual representa una ausencia de información de 14% de la población investigada. Esto, sumado a lo anterior, ratifica desconocimiento de la emisión de reportes de evaluación para públicos diversos interesados en la rendición de cuentas a través de los resultados del rendimiento escolar (como podrían ser autoridades gubernamentales u organismos certificadores). Por otra parte, en los demás estándares, las incidencias del valor 6 fueron menores de 8%.

#### *Consistencia interna y análisis de diferencias entre dos grupos*

La tabla 5 contiene el análisis de consistencia interna del instrumento. Podemos apreciar valores aceptables en todos los casos.

**Tabla 5.** Resultados alfa de Cronbach para estimar consistencia interna.

Alfa de Cronbach	Clasificación de estándares			
	Ética	Utilidad	Factibilidad	Precisión
	0.79	0.90	0.73	0.88

Aun cuando este estudio se propuso contar con una primera aproximación de las prácticas de evaluación en torno a estándares sin hacer precisiones sobre nivel educativo o nacionalidad, consideramos pertinente realizar una comparación de medias utilizando *Z-test* o prueba Z entre los grupos más numerosos: México y Colombia; esto, con la intención de saber si podría existir alguna diferencia entre sus percepciones y que esto diera pie a futuros estudios. Los datos revelaron que no hay diferencias significativas en general con un alfa de 0.05. Sin embargo, al efectuar el análisis segmentando por estándar, advertimos una diferencia significativa de las medias entre ambos grupos, en cuanto a factibilidad se refiere (2.82.> 1.96).

**Tabla 6.** Resultados comparación de medias.

	Ética		Utilidad		Factibilidad		Precisión		General	
	Prom.	Var.	Prom.	Var.	Prom.	Var.	Prom.	Var.	Prom.	Var.
Colombia	4.38	0.26	4.12	0.38	3.98	0.42	3.92	0.45	4.10	0.26
México	4.39	0.24	4.15	0.36	3.78	0.41	3.99	0.43	4.07	0.25
Prueba Z	-0.11		-0.31		2.82		-0.83		0.51	

N en Colombia = 183 y N en México = 134

## DISCUSIÓN

Un primer hallazgo se relaciona con los resultados en el estándar de precisión. Los puntajes obtenidos en la mayoría de los ítems que conforman este estándar mostraron una tendencia con una carga más negativa que positiva y una variabilidad mayor que cualquier otro grupo de afirmaciones. Esto permite inferir que en los ámbitos educativos tal vez existen pocos momentos para repensar en los procesos de evaluación en aras de mejorarlos y que la comunicación de los resultados, excepto para alumnos y padres, aparentemente no se encuentra dentro de la agenda de trabajo respecto a la evaluación. Este hallazgo es afín a la jerarquización de Cabra (2008), en la cual se percibe que las acciones relacionadas con la mejora del proceso de evaluación, así como el cuidado de su confiabilidad y validez, no son tan importantes como el estimar y hacer juicios sobre el aprendizaje.

Un segundo hallazgo a resaltar es la comunicación distante entre alumnos, padres y maestros respecto a los procesos de evaluación. Aunque es factible y deseable que esto suceda, al parecer la concepción de cómo y quién evalúa está desvinculada de la comunicación que podría establecerse sobre los procesos de evaluación. En el estudio conducido por Cabra (2008) se hace evidente que este aspecto suele no notarse por el alumno en el sentido global, sino sólo cuando a él le divulguen los resultados y le digan de dónde surgieron.

Un tercer hallazgo es que existe una diferencia entre dos grupos de profesores en el estándar de factibilidad. Esto podría indicarnos que a pesar de que en ambos países se tienen costumbres similares, además de compartir problemáticas educativas muy parecidas, se advierte una diferencia entre los grupos docentes sobre qué tan frecuentemente planean, dialogan y ejecutan procesos de evaluación.

Un último punto es que, de todos los estándares, el de ética es el que sostiene mejores resultados, ya que los profesores afirman que realizaron acciones de evaluación conduciéndose con respeto, consideración y cumplimento hacia los estudiantes. No obstante, la buena puntuación en ética es un estándar que demanda una serie de detalles que no siempre se cubren en la muestra encuestada, como se refleja en la baja puntuación sobre el uso de manuales para ejecutar procedimientos de evaluación, por ejemplo.

## CONCLUSIONES

Este proceso de indagación sobre cómo perciben profesores y directivos las prácticas de evaluación del aprendizaje valoradas desde la visión de los estándares ha permitido dar una primera vista sobre cuáles son las fortalezas y debilidades que experimentan los educadores en torno al tema, en diferentes niveles y contextos educativos. Los resultados hacen que se plantee una serie de interrogantes en materia de planeación, instrumentación, aplicación, análisis de datos y mejora continua que ameritan respuesta a partir de diversas preguntas que cuestionen la existencia de diferencias entre las acciones y decisiones tomadas por maestros, según contexto y nivel educativo en el cual laboran. Este tipo de investigaciones coadyuvará a orientar mejor los procesos de formación de profesores en materia de evaluación del aprendizaje, alineados a estándares internacionales, lo que permitirá cuidar el dominio de las técnicas para medir el aprendizaje y la correcta orientación de todo el proceso de emisión de juicios y alternativas para alcanzar las competencias u objetivos trazados, además de apoyar la realización de acciones que contribuirán no sólo a ejercer la evaluación del aprendizaje, sino construir sistemas de evaluación para el aprendizaje.

## REFERENCIAS BIBLIOGRÁFICAS

- Álvarez, J. (2001). *Evaluar para conocer, examinar para excluir*. Madrid, España: Morata.
- Banco Interamericano de Desarrollo (BID) (2011). *Los docentes, las escuelas y los aprendizajes escolares en América Latina: un estudio regional usando la base de datos del SERCE*. Recuperado el 15 de octubre de 2014 de <http://services.iadb.org/wmsfiles/products/Publications/35361923.pdf>
- Cabra, F. (2008). La calidad de la evaluación de los estudiantes: un análisis desde estándares profesionales. *Magis Revista Internacional de Investigación en Educación*. Recuperado de <http://www.redalyc.org/pdf/2810/281021687007.pdf>
- Clipa, O. y Duta, N. (2010). La percepción del alumnado acerca de la frecuencia de evaluación. *REIRE: Revista d'Innovació i Recerca en Educació*, vol. 3, núm. 2, pp. 53-67
- Escudero, T. (2013). Utilidad y uso de las evaluaciones. Un asunto relevante. *Revista de Evaluación Educativa REVALUE*, 2 (1). Recuperado el 15 de octubre de 2014 de <http://revalue.mx/revista/index.php/revalue/issue/current>
- Hernández, Fernández-Collado y Baptista (2010). *Metodología de la investigación* (4ta. ed.). Distrito Federal, México: McGraw-Hill.
- Martínez, F. (2013). El futuro de la evaluación educativa. *Sinéctica, revista electrónica de educación*, núm. 40. Recuperado de [http://www.sinectica.iteso.mx/?seccion=articulo&lang=es&id=591\\_el\\_futuro\\_de\\_la\\_evaluacion\\_educativa](http://www.sinectica.iteso.mx/?seccion=articulo&lang=es&id=591_el_futuro_de_la_evaluacion_educativa)
- Moreno, T. (2012). Evaluación para el aprendizaje. Perspectivas internacionales. *Revista de Evaluación Educativa REVALUE*, 1 (1). Recuperado el 1 de enero de 2014 de <http://revalue.mx/revista/index.php/revalue/issue/current>
- Kimball, S. & Milanowski, A. (2009). Examining Teacher Evaluation Validity and Leadership Decision Making Within a Standards-Based Evaluation System. *Educational Administration Quarterly February*, vol. 45, núm.1, pp. 34-70.
- Organización para la Cooperación y el Desarrollo Económicos (OCDE) (2014). *Panorama de la educación 2014-México*. Recuperado el 20 de diciembre de 2014 de <file:///localhost/E:/L00737965/Laboral/1 EGE/002 Bibliografía por temas/Generales de educación/Mexico-EAG2014-Country-Note-spanish.pdf>
- Peterson, K., Kelly, P. y Caskey, M. (2006, enero-marzo). Consideraciones éticas para los profesores en la evaluación de sus colegas. *Revista Mexicana de Investigación Educativa*, pp. 317-326.
- Pérez, R., Galán, A. y Quintanal, J. (2012). *Métodos y diseños de investigación en educación*. Madrid, España: UNED.
- Ravitch, D. (1996). *Estándares nacionales en educación*. Washington, EUA: Brookings Institution.
- Sanders, J. (1999). *General Background on the Joint Committee on Standards for Educational Evaluation*. Recuperado el 15 de julio de 2015 de <http://www.jcsee.org/wp-content/uploads/2009/09/JCGeneralBackground.pdf>
- Shapiro, E. (1973). Educational Evaluation: rethinking the criterion of competence. *The School Review* (pp. 523-549). The University of Chicago Press. Recuperado el 15 de enero de 2015 de <http://www.jstor.org/stable/1084424>
- UNESCO (2013). *Hacia un aprendizaje universal. Recomendaciones de la comisión especial sobre métricas de los aprendizajes*. Recuperado el 3 de enero de 2015 de <http://unesdoc.unesco.org/images/0022/002241/224102s.pdf>
- \_\_\_\_\_ (2014). *Guía del docente para la sensibilización en favor de una educación de calidad*. Recuperado el 12 de enero de 2015 de <http://unesdoc.unesco.org/images/0022/002299/229956s.pdf>
- Valenzuela, J. y Flores, M. (2012). *Fundamentos de investigación educativa*, vol. 2. Monterrey, Nuevo León, México: Editorial Digital del Tecnológico de Monterrey.
- Yarbrough, D., Shulha, L., Hopson, R., & Caruthers, F. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3ra. ed.). Thousand Oaks, CA: Sage.

---

### Acerca de los autores

#### Katherina Edith Gallardo Córdova

Doctora en Innovación y Tecnología Educativa por el Instituto Tecnológico y de Estudios Superiores de Monterrey. Profesora-investigadora en el Tecnológico de Monterrey, México.

#### Arcelia Ramos Monobe

Maestra en Educación con especialidad en Desarrollo Cognitivo. Profesora-tutora en el Tecnológico de Monterrey, México.

#### Dulce Camacho Gutiérrez

Maestra en Educación con especialidad en Desarrollo Cognitivo. Profesora-tutora en el Tecnológico de Monterrey, México.

#### María Eugenia Gil Rendón

Maestra en Ciencias de la Educación. Profesora-tutora en el Tecnológico de Monterrey, México.

---

Fecha de recepción del artículo: 01/02/2015  
Fecha de aceptación para su publicación: 15/07/2015